Graduate Texts in Mathematics

Patrick Morandi

Field and Galois Theory

Springer

Graduate Texts in Mathematics 167

Springer
New York
Berlin
Heidelberg
Barcelona
Budapest
Hong Kong
London
Milan
Paris
Santa Clara
Singapore
Tokyo

Graduate Texts in Mathematics

continued after index

Patrick Morandi

Field and Galois Theory

With 18 Illustrations

Springer

Patrick Morandi
Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003
USA

Mathematics Subject Classifications (1991): 12-01, 12F10, 12F20

Library of Congress Cataloging-in-Publication Data
Morandi, Patrick.
Field and Galois theory/Patrick Morandi.
p. cm. – (Graduate texts in mathematics ; 167)
Includes bibliographical references and index.
ISBN-13: 978-1-4612-8475-8 e-ISBN-13: 978-1-4612-4040-2
DOI: 10.1007/978-1-4612-4040-2
1. Algebraic fields. 2. Galois theory. I. Title. II. Series.
QA247.M67 1996
512′.3–dc20 96-13581

Printed on acid-free paper.

Softcover reprint of the hardcover 1st edition 1996

Production managed by Hal Henglein; manufacturing supervised by Joe Quatela.
Camera-ready copy prepared from the author's LaTeX files.

9 8 7 6 5 4 3 2 1

Preface

In the fall of 1990, I taught Math 581 at New Mexico State University for the first time. This course on field theory is the first semester of the year-long graduate algebra course here at NMSU. In the back of my mind, I thought it would be nice someday to write a book on field theory, one of my favorite mathematical subjects, and I wrote a crude form of lecture notes that semester. Those notes sat undisturbed for three years until late in 1993 when I finally made the decision to turn the notes into a book. The notes were greatly expanded and rewritten, and they were in a form sufficient to be used as the text for Math 581 when I taught it again in the fall of 1994.

Part of my desire to write a textbook was due to the nonstandard format of our graduate algebra sequence. The first semester of our sequence is field theory. Our graduate students generally pick up group and ring theory in a senior-level course prior to taking field theory. Since we start with field theory, we would have to jump into the middle of most graduate algebra textbooks. This can make reading the text difficult by not knowing what the author did before the field theory chapters. Therefore, a book devoted to field theory is desirable for us as a text. While there are a number of field theory books around, most of these were less complete than I wanted. For example, Artin's wonderful book [1] barely addresses separability and does not deal with infinite extensions. I wanted to have a book containing most everything I learned and enjoyed about field theory.

This leads to another reason why I wanted to write this book. There are a number of topics I wanted to have in a single reference source. For instance, most books do not go into the interesting details about discriminants and

how to calculate them. There are many versions of discriminants in different fields of algebra. I wanted to address a number of notions of discriminant and give relations between them. For another example, I wanted to discuss both the calculation of the Galois group of a polynomial of degree 3 or 4, which is usually done in Galois theory books, and discuss in detail the calculation of the roots of the polynomial, which is usually not done. I feel it is instructive to exhibit the splitting field of a quartic as the top of a tower of simple radical extensions to stress the connection with solvability of the Galois group. Finally, I wanted a book that does not stop at Galois theory but discusses non-algebraic extensions, especially the extensions that arise in algebraic geometry. The theory of finitely generated extensions makes use of Galois theory and at the same time leads to connections between algebra, analysis, and topology. Such connections are becoming increasingly important in mathematical research, so students should see them early.

The approach I take to Galois theory is roughly that of Artin. This approach is how I first learned the subject, and so it is natural that I feel it is the best way to teach Galois theory. While I agree that the fundamental theorem is the highlight of Galois theory, I feel strongly that the concepts of normality and separability are vital in their own right and not just technical details needed to prove the fundamental theorem. It is due to this feeling that I have followed Artin in discussing normality and separability before the fundamental theorem, and why the sections on these topics are quite long. To help justify this, I point out that results in these sections are cited in subsequent chapters more than is the fundamental theorem.

This book is divided into five chapters, along with five appendices for background material. The first chapter develops the machinery of Galois theory, ending with the fundamental theorem and some of its most immediate consequences. One of these consequences, a proof of the fundamental theorem of algebra, is a beautiful application of Galois theory and the Sylow theorems of group theory. This proof made a big impression on me when I first saw it, and it helped me appreciate the Sylow theorems.

Chapter II applies Galois theory to the study of certain field extensions, including those Galois extensions with a cyclic or Abelian Galois group. This chapter takes a diversion in Section 10. The classical proof of the Hilbert theorem 90 leads naturally into group cohomology. While I believe in giving students glimpses into more advanced topics, perhaps this section appears in this book more because of my appreciation for cohomology. As someone who does research in division algebras, I have seen cohomology used to prove many important theorems, so I felt it was a topic worth having in this book.

In Chapter III, some of the most famous mathematical problems of antiquity are presented and answered by using Galois theory. The main questions of ruler and compass constructions left unanswered by the ancient Greeks, such as whether an arbitrary angle can be trisected, are resolved. We combine analytic and algebraic arguments to prove the transcendence of π and

e. Formulas for the roots of cubic and quartic polynomials, discovered in the sixteenth century, are given, and we prove that no algebraic formula exists for the roots of an arbitrary polynomial of degree 5 or larger. The question of solvability of polynomials led Galois to develop what we now call Galois theory and in so doing also developed group theory. This work of Galois can be thought of as the birth of abstract algebra and opened the door to many beautiful theories.

The theory of algebraic extensions does not end with finite extensions. Chapter IV discusses infinite Galois extensions and presents some important examples. In order to prove an analog of the fundamental theorem for infinite extensions, we need to put a topology on the Galois group. It is through this topology that we can determine which subgroups show up in the correspondence between subextensions of a Galois extension and subgroups of the Galois group. This marks just one of the many places in algebra where use of topology leads to new insights.

The final chapter of this book discusses nonalgebraic extensions. The first two sections develop the main tools for working with transcendental extensions: the notion of a transcendence basis and the concept of linear disjointness. The latter topic, among other things, allows us to extend to arbitrary extensions the idea of separability. The remaining sections of this chapter introduce some of the most basic ideas of algebraic geometry and show the connections between algebraic geometry and field theory, notably the theory of finitely generated nonalgebraic extensions. It is the aim of these sections to show how field theory can be used to give geometric information, and vice versa. In particular, we show how the dimension of an algebraic variety can be calculated from knowledge of the field of rational functions on the variety.

The five appendices give what I hope is the necessary background in set theory, group theory, ring theory, vector space theory, and topology that readers of this book need but in which they may be partially deficient. These appendices are occasionally sketchy in details. Some results are proven and others are quoted as references. Their purpose is not to serve as a text for these topics but rather to help students fill holes in their background. Exercises are given to help to deepen the understanding of these ideas.

Two things I wanted this book to have were lots of examples and lots of exercises. I hope I have succeeded in both. One complaint I have with some field theory books is a dearth of examples. Galois theory is not an easy subject to learn. I have found that students often finish a course in Galois theory without having a good feel for what a Galois extension is. They need to see many examples in order to really understand the theory. Some of the examples in this book are quite simple, while others are fairly complicated. I see no use in giving only trivial examples when some of the interesting mathematics can only be gleaned from looking at more intricate examples. For this reason, I put into this book a few fairly complicated and nonstandard examples. The time involved in understanding these examples

will be time well spent. The same can be said about working the exercises. It is impossible to learn any mathematical subject merely by reading text. Field theory is no exception. The exercises vary in difficulty from quite simple to very difficult. I have not given any indication of which are the hardest problems since people can disagree on whether a problem is difficult or not. Nor have I ordered the problems in any way, other than trying to place a problem in a section whose ideas are needed to work the problem. Occasionally, I have given a series of problems on a certain theme, and these naturally are in order. I have tried not to place crucial theorems as exercises, although there are a number of times that a step in a proof is given as an exercise. I hope this does not decrease the clarity of the exposition but instead improves it by eliminating some simple but tedious steps.

Thanks to many people need to be given. Certainly, authors of previously written field theory books need to be thanked; my exposition has been influenced by reading these books. Adrian Wadsworth taught me field theory, and his teaching influenced both the style and content of this book. I hope this book is worthy of that teaching. I would also like to thank the colleagues with whom I have discussed matters concerning this book. Al Sethuraman read preliminary versions of this book and put up with my asking too many questions, Irena Swanson taught Math 581 in fall 1995 using it, and David Leep gave me some good suggestions. I must also thank the students of NMSU who put up with mistake-riddled early versions of this book while trying to learn field theory. Finally, I would like to thank the employees at TCI Software, the creators of Scientific Workplace. They gave me help on various aspects of the preparation of this book, which was typed in LaTeX using Scientific Workplace.

April 1996 Pat Morandi
Las Cruces, New Mexico

Notes to the Reader

The prerequisites for this book are a working knowledge of ring theory, including polynomial rings, unique factorization domains, and maximal ideals; some group theory, especially finite group theory; vector space theory over an arbitrary field, primarily existence of bases for finite dimensional vector spaces, and dimension. Some point set topology is used in Sections 17 and 21. However, these sections can be read without worrying about the topological notions. Profinite groups arise in Section 18 and tensor products arise in Section 20. If the reader is unfamiliar with any of these topics, as mentioned in the Preface there are five appendices at the end of the book that cover these concepts to the depth that is needed. Especially important is Appendix A. Facts about polynomial rings are assumed right away in Section 1, so the reader should peruse Appendix A to see if the material is familiar.

The numbering scheme in this book is relatively simple. Sections are numbered independently of the chapters. A theorem number of 3.5 means that the theorem appears in Section 3. Propositions, definitions, etc., are numbered similarly and in sequence with each other. Equation numbering follows the same scheme. A problem referred to in the section that it appears will be labeled such as Problem 4. A problem from another section will be numbered as are theorems; Problem 13.3 is Problem 3 of Section 13. This numbering scheme starts over in each appendix. For instance, Theorem 2.3 in Appendix A is the third numbered item in the second section of Appendix A.

Definitions in this book are given in two ways. Many definitions, including all of the most important ones, are spelled out formally and assigned a

number. Other definitions and some terminology are given in the body of the text and are emphasized by italic text. If this makes it hard for a reader to find a definition, the index at the end of the book will solve this problem.

There are a number of references at the end of the book, and these are cited occasionally throughout the book. These other works are given mainly to allow the reader the opportunity to see another approach to parts of field theory or a more in-depth exposition of a topic. In an attempt to make this book mostly self-contained, substantial results are not left to be found in another source. Some of the theorems are attributed to a person or persons, although most are not. Apologies are made to anyone, living or dead, whose contribution to field theory has not been acknowledged.

Notation in this book is mostly standard. For example, the subset relation is denoted by $\subseteq$ and proper subset by $\subset$. If B is a subset of A, then the set difference $\{x : x \in A, x \notin B\}$ is denoted by $A - B$. If I is an ideal in a ring R, the coset $r + I$ is often denoted by $\overline{r}$. Most of the notation used is given in the List of Symbols section. In that section, each symbol is given a page reference where the symbol can be found, often with definition.

Contents

List of Symbols

Listed here are most of the symbols used in the text, along with the meaning and a page reference for each symbol.

Symbol	Meaning	Page
$\subseteq$	subset	1
K/F	field extension $F \subseteq K$	1
$[K : F]$	degree of field extension	2
$\mathbb{N}$	natural numbers	2
$\mathbb{Z}$	integers	2
$\mathbb{Q}$	rational numbers	2
$\mathbb{R}$	real numbers	2
$\mathbb{C}$	complex numbers	2
$\mathbb{F}_p$	integers mod p	2
$\{a : \mathcal{P}(a)\}$	set builder notation	3
ev_a	evaluation homomorphism	5
$F[\alpha], F(\alpha)$	ring and field generated by F and α	5
$F[\alpha_1, \dots, \alpha_n]$	ring generated by F and $\alpha_1, \dots \alpha_n$	5
$F(\alpha_1, \dots, \alpha_n)$	field generated by F and $\alpha_1, \dots, \alpha_n$	5
$F(X)$	field generated by F and X	5
$\min(F, \alpha)$	minimal polynomial of α over F	6
$\ker(\varphi)$	kernel of φ	7
$\deg(f(x))$	degree of $f(x)$	8
$\gcd(f(x), g(x))$	greatest common divisor	8

I
Galois Theory

In this chapter, we develop the machinery of Galois theory. The first four sections constitute the technical heart of Galois theory, and Section 5 presents the fundamental theorem and some consequences. As an application, we give a proof of the fundamental theorem of algebra using Galois theory and the Sylow theorems of group theory.

The main idea of Galois theory is to associate a group, the Galois group, to a field extension. We can then turn field theory problems into group theory problems. Since the Galois group of a finite dimensional extension is finite, we can utilize the numerical information about finite groups to help investigate such field extensions. It turns out that field theory is the right context for solving some of the famous classical problems that stumped mathematicians for centuries. As an application of field theory, in Chapter III we give proofs of the famous impossibilities of certain ruler and compass constructions, and we determine why roots of polynomials of degree 5 or greater need not be given by formulas involving field operations and extraction of roots.

1 Field Extensions

In this section, we begin the study of field theory. Consequently, there are a number of definitions in this section, although there are also a large number of examples intended to help the reader with the concepts. We point out now that we take a basic knowledge of ring theory and vector space theory

for granted. For instance, we use the dimension of a finite dimensional vector space frequently, and we use the theory of polynomial rings in one variable over a field equally often. Any reader who is unfamiliar with a fact used in this book is recommended to peruse the appendices; they contain most of the background a reader will need but may not have.

While field theory is of course concerned with the study of fields, the study of field theory primarily investigates field extensions. In fact, the classical problems of ruler and compass constructions and the solvability of polynomial equations were answered by analyzing appropriate field extensions, and we answer these problems in Chapter III in this way. While it may seem unusual to some readers to consider pairs of fields, we point out that much of group theory and ring theory is concerned with group extensions and ring extensions, respectively.

Recall that a field is a commutative ring with identity such that the nonzero elements form a group under multiplication. If $F \subseteq K$ are fields, then K is called a *field extension* of F. We will refer to the pair $F \subseteq K$ as the field extension K/F and to F as the *base field*. We make K into an F-vector space by defining scalar multiplication for $\alpha \in F$ and $a \in K$ as $\alpha \cdot a = \alpha a$, the multiplication of α and a in K. We write $[K : F]$ for the dimension of K as an F-vector space. This dimension is called the *degree* of K/F. If $[K : F] < \infty$, then K is called a *finite extension* of F. Otherwise K is an *infinite extension* of F. Most of this chapter will deal with finite field extensions, although in a few places we will need to work with extensions of any degree.

Example 1.1 In order to give examples of field extensions, we first need examples of fields. In this book, the fields of *rational numbers*, *real numbers*, and *complex numbers* will be denoted $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, respectively. The field $\mathbb{Z}/p\mathbb{Z}$ of *integers mod p* will be denoted $\mathbb{F}_p$. The fields $\mathbb{Q}$ and $\mathbb{F}_p$ will appear often as the base field of examples. Finite field extensions of $\mathbb{Q}$ are called *algebraic number fields* and are one of the objects of study in algebraic number theory.

Example 1.2 Let k be a field and let x be a variable. The rational function field $k(x)$ is the quotient field of the polynomial ring $k[x]$; that is, $k(x)$ consists of all quotients $f(x)/g(x)$ of polynomials with $g(x) \neq 0$. Similarly, if $x_1, \dots, x_n$ are independent variables, then the field $k(x_1, \dots, x_n)$ of rational functions in the x_i is the quotient field of the polynomial ring $k[x_1, \dots, x_n]$ of polynomials in n variables, so it consists of all quotients $f(x_1, \dots, x_n)/g(x_1, \dots, x_n)$ of polynomials in the x_i with $g \neq 0$. Field extensions of a rational function field arise frequently in algebraic geometry and in the theory of division rings. We will work with rational function fields frequently.

Example 1.3 Let k be a field and let $k((x))$ be the set of all formal generalized power series in x with coefficients in k; that is, the elements of $k((x))$ are formal infinite sums $\sum_{n=n_0}^{\infty} a_n x^n$ with $n_0 \in \mathbb{Z}$ and each $a_n \in k$. We define addition and multiplication on $k((x))$ by

$$\sum_n^{\infty} a_n x^n + \sum_n^{\infty} b_n x^n = \sum_n^{\infty} (a_n + b_n) x^n$$

and

$$\sum_{n=n_0}^{\infty} a_n x^n \cdot \sum_{n=n_1}^{\infty} b_n x^n = \sum_{n=n_0+n_1}^{\infty} \left(\sum_{k=n_0}^{n-n_1} a_n b_{n-k} \right) x^n.$$

A straightforward calculation shows that $k((x))$ is a commutative ring with identity. Moreover, we can show that $k((x))$ is a field. If $f = \sum_{n=n_0}^{\infty} a_n x^n$ is a nonzero element of $k((x))$, we need to produce an inverse for f. Suppose that we have written the series so that a_{n_0} is the first nonzero coefficient. By multiplying by $a_{n_0}^{-1} x^{-n_0}$, to find an inverse for f it suffices to assume that $n_0 = 0$ and $a_{n_0} = 1$. We can find the coefficients b_n of the inverse $\sum_{n=0}^{\infty} b_n x^n$ to f by recursion. To have $\sum_{n=0}^{\infty} a_n x^n \cdot \sum_{n=0}^{\infty} b_n x^n = 1$, we need $b_0 = 1$ since $a_0 = 1$. For $n > 0$, the coefficient of x^n is

$$b_n a_0 + b_{n-1} a_1 + \cdots + b_0 a_n = 0,$$

so if we have determined $b_0, \ldots, b_{n-1}$, then we determine b_n from the equation $b_n = -\sum_{k=1}^{n} b_{n-k} a_k$. By setting g to be the series with coefficients b_n determined by this information, our computations yield $fg = 1$. Thus, $k((x))$ is a field. The rational function field $k(x)$ is naturally isomorphic to a subfield of $k((x))$. In algebra, the field $k((x))$ is often called the field of Laurent series over k, although this terminology is different from that used in complex analysis.

We now give some examples of field extensions.

Example 1.4 The extension $\mathbb{C}/\mathbb{R}$ is a finite extension since $[\mathbb{C} : \mathbb{R}] = 2$. A basis for $\mathbb{C}$ as an $\mathbb{R}$-vector space is $\{1, i\}$. As an extension of $\mathbb{Q}$, both $\mathbb{C}$ and $\mathbb{R}$ are infinite extensions. If $a \in \mathbb{C}$, let

$$\mathbb{Q}(a) = \left\{ \frac{\sum_i \alpha_i a^i}{\sum_i \beta_i a^i} : \alpha_i, \beta_i \in \mathbb{Q}, \sum_i \beta_i a^i \neq 0 \right\}.$$

We shall see in Proposition 1.8 that $\mathbb{Q}(a)$ is a field extension of $\mathbb{Q}$. The degree of $\mathbb{Q}(a)/\mathbb{Q}$ can be either finite or infinite depending on a. For instance, if $a = \sqrt{-1}$ or $a = \exp(2\pi i/3)$, then $[\mathbb{Q}(a) : \mathbb{Q}] = 2$. These equalities are consequences of Proposition 1.15. On the other hand, we prove in Section 14 that $[\mathbb{Q}(\pi) : \mathbb{Q}] = \infty$.

Example 1.5 If k is a field, let $K = k(t)$ be the field of rational functions in t over k. If f is a nonzero element of K, then we can use the construction of $\mathbb{Q}(a)$ in the previous example. Let $F = k(f)$ be the set of all rational functions in f; that is,

$$F = \left\{ \frac{\sum_{i=0}^{n} a_i f^i}{\sum_{j=0}^{m} b_j f^j} : a_i, b_j \in k \text{ and } \sum_{j=0}^{m} b_j f^j \neq 0 \right\}.$$

If $f(t) = t^2$, then K/F is an extension of degree 2; a basis for K is $\{1, t\}$. In Example 1.17, we shall see that K/F is a finite extension provided that f is not a constant, and in Chapter V we shall prove Lüroth's theorem, which states that every field L with $k \subseteq L \subseteq K$ is of the form $L = k(f)$ for some $f \in K$.

Example 1.6 Let $p(t) = t^3 - 2 \in \mathbb{Q}[t]$. Then $p(t)$ is irreducible over $\mathbb{Q}$ by the rational root test. Then the ideal $(p(t))$ generated by $p(t)$ in $\mathbb{Q}[t]$ is maximal; hence, $K = \mathbb{Q}[t]/(p(t))$ is a field. The set of cosets $\{a + (p(t)) : a \in \mathbb{Q}\}$ can be seen to be a field isomorphic to $\mathbb{Q}$ under the map $a \mapsto a + (p(t))$. We view the field $\mathbb{Q}[t]/(p(t))$ as an extension field of $\mathbb{Q}$ by thinking of $\mathbb{Q}$ as this isomorphic subfield. If $f(t) \in \mathbb{Q}[t]$, then by the division algorithm, $f(t) = q(t)p(t) + r(t)$ with $r(t) = 0$ or $\deg(r) < \deg(p) = 3$. Moreover, $f(t)$ and $r(t)$ generate the same coset in $\mathbb{Q}[t]/(p(t))$. What this means is that any element of K has a unique representation in the form $a + bt + ct^2 + (p(t))$ for some $a, b, c \in \mathbb{Q}$. Therefore, the cosets $1 + (p(t))$, $t + (p(t))$, and $t^2 + (p(t))$ form a basis for K over $\mathbb{Q}$, so $[K : \mathbb{Q}] = 3$. Let $a = t + (p(t))$. Then

$$a^3 - 2 = t^3 + (p(t)) - (2 + (p(t)) = t^3 - 2 + (p(t)) = 0.$$

The element a is then a root of $x^3 - 2$ in K. Note that we used the identification of $\mathbb{Q}$ as a subfield in this calculation.

If instead of $t^3 - 2$ we had started with any irreducible polynomial of degree n over $\mathbb{Q}$, we would obtain a field extension of $\mathbb{Q}$ of degree n that contains a root of the polynomial. We will use this idea in Section 3 to prove the existence of fields that contain roots of polynomials.

Generators of fields

In order to study the roots of a polynomial over a field F, we will consider a minimal field extension of F that contains all the roots of the polynomial. In intuitive terms, we want this field to be generated by F and the roots. We need to make this more precise.

Definition 1.7 *Let K be a field extension of F. If X is a subset of K, then the ring $F[X]$ generated by F and X is the intersection of all subrings of K that contain F and X. The field $F(X)$ generated by F and X is the intersection of all subfields of K that contain F and X. If $X = \{a_1, \ldots, a_n\}$*

is finite, we will write $F[X] = F[a_1, \ldots, a_n]$ *and* $F(X) = F(a_1, \ldots, a_n)$. *If* X *is finite, we call the field* $F(X)$ *a finitely generated extension of* F.

It is a simple exercise to show that an intersection of subfields or subrings of a field is again a subfield or subring, respectively. From this definition, it follows that $F(X)$ is the smallest subfield with respect to inclusion of K that contains F and X. We can give more concrete descriptions of $F[X]$ and $F(X)$. Let K be a field extension of F and let $a \in K$. The *evaluation homomorphism* ev_a is the map $\mathrm{ev}_a : F[x] \to K$ defined by $\mathrm{ev}_a(\sum_i \alpha_i x^i) = \sum_i \alpha_i a^i$. We denote $\mathrm{ev}_a(f(x))$ by $f(a)$. It is straightforward (see Problem 3) to show that ev_a is both a ring and an F-vector space homomorphism. We use this notion to see what it means for a field to be generated by a set of elements. We start with the easiest case, when K is generated over F by a single element.

Proposition 1.8 *Let* K *be a field extension of* F *and let* $a \in K$. *Then*

$$F[a] = \{f(a) : f(x) \in F[x]\}$$

and

$$F(a) = \{f(a)/g(a) : f, g \in F[x], g(a) \neq 0\}.$$

Moreover, $F(a)$ *is the quotient field of* $F[a]$.

Proof. The evaluation map $\mathrm{ev}_a : F[x] \to K$ has image $\{f(a) : f \in F[x]\}$, so this set is a subring of K. If R is a subring of K that contains F and a, then $f(a) \in R$ for any $f(x) \in F[x]$ by closure of addition and multiplication. Therefore, $\{f(a) : f(x) \in F[x]\}$ is contained in all subrings of K that contain F and a. Therefore, $F[a] = \{f(a) : f(x) \in F[x]\}$. The quotient field of $F[a]$ is then the set $\{f(a)/g(a) : f, g \in F[x], g(a) \neq 0\}$. It clearly is contained in any subfield of K that contains $F[a]$; hence, it is equal to $F(a)$. □

The notation $F[a]$ and $F(a)$ is consistent with the notation $F[x]$ and $F(x)$ for the ring of polynomials and field of rational functions over F, as the description of $F[a]$ and $F(a)$ shows.

By similar arguments, we can describe the ring $F[a_1, \ldots, a_n]$ and field $F(a_1, \ldots, a_n)$ generated by F and $a_1, \ldots, a_n$. The proof of the following proposition is not much different from the proof of Proposition 1.8, so it is left to Problem 4.

Proposition 1.9 *Let* K *be a field extension of* F *and let* $a_1, \ldots, a_n \in K$. *Then*

$$F[a_1, \ldots, a_n] = \{f(a_1, \ldots, a_n) : f \in F[x_1, \ldots, x_n]\}$$

and

$$F(a_1, \ldots, a_n) = \left\{ \frac{f(a_1, \ldots, a_n)}{g(a_1, \ldots, a_n)} : f, g \in F[x_1, \ldots, x_n],\ g(a_1, \ldots, a_n) \neq 0 \right\},$$

so $F(a_1, \ldots, a_n)$ *is the quotient field of* $F[a_1, \ldots, a_n]$.

For arbitrary subsets X of K we can describe the field $F(X)$ in terms of finite subsets of X. This description is often convenient for turning questions about field extensions into questions about finitely generated field extensions.

Proposition 1.10 *Let* K *be a field extension of* F *and let* X *be a subset of* K. *If* $\alpha \in F(X)$, *then* $\alpha \in F(a_1, \ldots, a_n)$ *for some* $a_1, \ldots, a_n \in X$. *Therefore,*

$$F(X) = \bigcup \{F(a_1, \ldots, a_n) : a_1, \ldots, a_n \in X\},$$

where the union is over all finite subsets of X.

Proof. Each field $F(a_1, \ldots, a_n)$ with the $a_i \in X$ is contained in $F(X)$; hence, $\bigcup\{F(a_1, \ldots, a_n) : a_i \in X\} \subseteq F(X)$. This union contains F and X, so if it is a field, then it is equal to $F(X)$, since $F(X)$ is the smallest subfield of K containing F and X. To show that this union is a field, let $\alpha, \beta \in \bigcup\{F(a_1, \ldots, a_n) : a_i \in X\}$. Then there are $a_i, b_i \in X$ with $\alpha \in F(a_1, \ldots, a_n)$ and $\beta \in F(b_1, \ldots, b_m)$. Then both α and β are contained in $F(a_1, \ldots, a_n, b_1, \ldots, b_m)$, so $\alpha \pm \beta$, $\alpha\beta$, and α/β (if $\beta \neq 0$) all lie in $\bigcup\{F(a_1, \ldots, a_n) : a_i \in X\}$. This union is then a field, so $F(X) = \bigcup\{F(a_1, \ldots, a_n) : a_i \in X\}$. □

In this chapter, our interest will be in those field extensions K/F for which any $a \in K$ satisfies a polynomial equation over F. We give this idea a formal definition.

Definition 1.11 *If* K *is a field extension of* F, *then an element* $\alpha \in K$ *is algebraic over* F *if there is a nonzero polynomial* $f(x) \in F[x]$ *with* $f(\alpha) = 0$. *If* α *is not algebraic over* F, *then* α *is said to be transcendental over* F. *If every element of* K *is algebraic over* F, *then* K *is said to be algebraic over* F, *and* K/F *is called an algebraic extension.*

Definition 1.12 *If* α *is algebraic over a field* F, *the minimal polynomial of* α *over* F *is the monic polynomial* $p(x)$ *of least degree in* $F[x]$ *for which* $p(\alpha) = 0$; *it is denoted* $\min(F, \alpha)$. *Equivalently,* $\min(F, \alpha)$ *is the monic generator* $p(x)$ *of the kernel of the evaluation homomorphism* ev_α.

Example 1.13 The complex number $i = \sqrt{-1}$ is algebraic over $\mathbb{Q}$, since $i^2 + 1 = 0$. If $r \in \mathbb{Q}$, then $a = \sqrt[n]{r}$ is algebraic over $\mathbb{Q}$, since a is a root of $x^n - r$. If $\omega = e^{2\pi i/n} = \cos(2\pi/n) + i\sin(2\pi/n)$, then $\omega^n - 1 = 0$, so ω is algebraic over $\mathbb{Q}$. Note that $\min(\mathbb{Q}, i) = x^2 + 1 = \min(\mathbb{R}, i)$ but $\min(\mathbb{C}, i) = x - i$. Therefore, the minimal polynomial of an element depends on the base field, as does whether the element is algebraic or transcendental. The determination of $\min(\mathbb{Q}, \omega)$ is nontrivial and will be done in Section 7.

Example 1.14 In 1873, Hermite proved that e is transcendental over $\mathbb{Q}$, and 9 years later, Lindemann proved that π is transcendental over $\mathbb{Q}$. However, π is algebraic over $\mathbb{Q}(\pi)$, since π is a root of the polynomial $x - \pi \in \mathbb{Q}(\pi)[x]$. It is unknown if e is transcendental over $\mathbb{Q}(\pi)$. We will prove in Section 14 that π and e are transcendental over $\mathbb{Q}$.

To work with algebraic extensions, we need some tools at our disposal. The minimal polynomial of an element and the degree of a field extension are two of the most basic tools we shall use. The following proposition gives a relation between these objects.

Proposition 1.15 *Let K be a field extension of F and let $\alpha \in K$ be algebraic over F.*

1. *The polynomial* $\min(F, \alpha)$ *is irreducible over F.*
2. *If $g(x) \in F[x]$, then $g(\alpha) = 0$ if and only if* $\min(F, \alpha)$ *divides $g(x)$.*
3. *If $n = \deg(\min(F, \alpha))$, then the elements $1, \alpha, \dots, \alpha^{n-1}$ form a basis for $F(\alpha)$ over F, so $[F(\alpha) : F] = \deg(\min(F, \alpha)) < \infty$. Moreover, $F(\alpha) = F[\alpha]$.*

Proof. If $p(x) = \min(F, \alpha)$, then $F[x]/(p(x)) \cong F[\alpha]$ is an integral domain. Therefore, $(p(x))$ is a prime ideal, so $p(x)$ is irreducible. To prove statement 2, if $g(x) \in F[x]$ with $g(\alpha) = 0$, then $g(x) \in \ker(\mathrm{ev}_\alpha)$. But this kernel is the ideal generated by $p(x)$, so $p(x)$ divides $g(x)$. For statement 3, we first prove that $F[\alpha] = F(\alpha)$. To see this, note that $F[\alpha]$ is the image of the evaluation map ev_α. The kernel of ev_α is a prime ideal since ev_α maps $F[x]$ into an integral domain. However, $F[x]$ is a principal ideal domain, so every nonzero prime ideal of $F[x]$ is maximal. Thus, $\ker(\mathrm{ev}_a)$ is maximal, so $F[\alpha] \cong F[x]/\ker(\mathrm{ev}_\alpha)$ is a field. Consequently, $F[\alpha] = F(\alpha)$. To finish the proof of statement 3, let $n = \deg(p(x))$. If $b \in F(\alpha)$, then $b = g(\alpha)$ for some $g(x) \in F[x]$. By the division algorithm, $g(x) = q(x)p(x) + r(x)$, where $r(x) = 0$ or $\deg(r) < n$. Thus, $b = g(\alpha) = r(\alpha)$. Since $r(\alpha)$ is an F-linear combination of $1, \alpha, \dots, \alpha^{n-1}$, we see that $1, \alpha, \dots, \alpha^{n-1}$ span $F(\alpha)$ as an F-vector space. If $\sum_{i=0}^{n-1} a_i \alpha^i = 0$, then $f(x) = \sum_{i=0}^{n-1} a_i x^i$ is divisible by $p(x)$, so $f(x) = 0$, or else f is divisible by a polynomial of larger degree than itself. Thus, $1, \alpha, \dots, \alpha^{n-1}$ is a basis for $F(\alpha)$ over F. □

Example 1.16 The element $\sqrt[3]{2}$ satisfies the polynomial $x^3 - 2$ over $\mathbb{Q}$, which is irreducible by the Eisenstein criterion, so $x^3 - 2$ is the minimal polynomial of $\sqrt[3]{2}$ over $\mathbb{Q}$. Thus, $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$. If p is a prime, then $x^n - p$ is irreducible over $\mathbb{Q}$, again by Eisenstein, so $[\mathbb{Q}(\sqrt[n]{p}) : \mathbb{Q}] = n$. The complex number $\omega = \cos(2\pi/3) + i\sin(2\pi/3)$ satisfies $x^3 - 1$ over $\mathbb{Q}$. This factors as $x^3 - 1 = (x-1)(x^2 + x + 1)$. The second factor has ω as a root and is irreducible since it has no rational root; hence, it is the minimal polynomial of ω over $\mathbb{Q}$. Consequently, $[\mathbb{Q}(\omega) : \mathbb{Q}] = 2$.

Let p be a prime and let $\rho = \exp(2\pi i/p) = \cos(2\pi/p) + i\sin(2\pi/p)$. Then ρ satisfies the polynomial $x^p - 1 = (x-1)(x^{p-1} + x^{p-2} + \cdots + x + 1)$. Since $\rho \neq 1$, it satisfies the polynomial $x^{p-1} + x^{p-2} + \cdots + x + 1$. Moreover, this polynomial is irreducible over $\mathbb{Q}$ (see Problem 22b); hence, it is the minimal polynomial of ρ over $\mathbb{Q}$.

Example 1.17 Here is a very nice, nontrivial example of a finite field extension. Let k be a field and let $K = k(t)$ be the field of rational functions in t over k. Let $u \in K$ with $u \notin k$. Write $u = f(t)/g(t)$ with $f, g \in k[t]$ and $\gcd(f(t), g(t)) = 1$, and let $F = k(u)$. We claim that

$$[K : F] = \max\{\deg(f(t), \deg(g(t))\},$$

which will show that K/F is a finite extension. To see this, first note that $K = F(t)$. By using Proposition 1.15, we need to determine the minimal polynomial of t over F to determine $[K : F]$. Consider the polynomial $p(x) = ug(x) - f(x) \in F[x]$. Then t is a root of $p(x)$. Therefore, t is algebraic over F, and so $[K : F] < \infty$ as $K = F(t)$. Say $f(t) = \sum_{i=0}^n a_i t^i$ and $g(t) = \sum_{i=0}^m b_i t^i$. First note that $\deg(p(x)) = \max\{\deg(f(t), \deg(g(t))\}$. If this were false, then the only way this could happen would be if $m = n$ and the coefficient of x^n in $p(x)$ were zero. But this coefficient is $ub_n - a_n$, which is nonzero since $u \notin k$. We now show that $p(x)$ is irreducible over F, which will verify that $[K : F] = \max\{n, m\}$. We do this by viewing $p(x)$ in two ways. The element u is not algebraic over k, otherwise $[K : k] = [K : F] \cdot [F : k] < \infty$, which is false. Therefore, u is transcendental over k, so $k[u] \cong k[x]$. Viewing p as a polynomial in u, we have $p \in k[x][u] \subseteq k(x)[u]$, and p has degree 1 in u. Therefore, p is irreducible over $k(x)$. Moreover, since $\gcd(f(t), g(t)) = 1$, the polynomial p is primitive in $k[x][u]$. Therefore, p is irreducible over $k[x]$. We have $p \in k[u][x] = k[x][u]$ (think about this!), so p is irreducible over $k[u]$, as a polynomial in x. Therefore, p is irreducible over $k(u) = F$, which shows that p is the minimal polynomial of u over F, by Proposition 1.15. Therefore, we have $[K : F] = \max\{\deg(f(t), \deg(g(t))\}$, as desired.

Example 1.18 Let K be a finitely generated extension of F, and suppose that $K = F(a_1, \ldots, a_n)$. We can break up the extension K/F into a collection of subextensions that are easier to analyze. Let $L_i = F(a_1, \ldots, a_i)$, and set $L_0 = F$. Then we have a chain of fields

$$F = L_0 \subseteq L_1 \subseteq L_2 \subseteq \cdots \subseteq L_n = K$$

with $L_{i+1} = L_i(a_{i+1})$. Therefore, we can break up the extension K/F into a series of subextensions L_{i+1}/L_i, each generated by a single element. Results such as Proposition 1.15 will help to study the extensions L_{i+1}/L_i. To make this idea of decomposing K/F into these subextensions useful, we will need to have *transitivity* results that tell us how to translate information

about subextensions to the full extension K/F. We will prove a number of transitivity results in this book. We prove two below, one dealing with field degrees and the other about the property of being algebraic.

Recall that the field K is finitely generated as a field over F if $K = F(a_1, \dots, a_n)$ for some $a_i \in K$. This is not the same as being finitely generated as a vector space or as a ring. The field K is finitely generated as an F-vector space if and only if $[K : F] < \infty$, and K is finitely generated as a ring over F if $K = F[\alpha_1, \dots, \alpha_n]$ for some $\alpha_i \in K$.

Lemma 1.19 *If K is a finite extension of F, then K is algebraic and finitely generated over F.*

Proof. Suppose that $\alpha_1, \dots, \alpha_n$ is a basis for K over F. Then every element of K is of the form $\sum_i a_i \alpha_i$ with $a_i \in F$, so certainly we have $K = F(\alpha_1, \dots, \alpha_n)$; thus, K is finitely generated over F. If $a \in K$, then $\{1, a, \dots, a^n\}$ is dependent over F, since $[K : F] = n$. Thus, there are $\beta_i \in F$, not all zero, with $\sum_i \beta_i a^i = 0$. If $f(x) = \sum_i \beta_i x^i$, then $f(x) \in F[x]$ and $f(a) = 0$. Therefore, a is algebraic over F, and so K is algebraic over F. □

The converse of this lemma is also true. In order to give a proof of the converse, we need the following property of degrees. The degree of a field extension is the most basic invariant of an extension. It is therefore important to have some information about this degree. We will use the following transitivity result frequently.

Proposition 1.20 *Let $F \subseteq L \subseteq K$ be fields. Then*

$$[K : F] = [K : L] \cdot [L : F].$$

Proof. Let $\{ a_i : i \in I \}$ be a basis for L/F, and let $\{ b_j : j \in J \}$ be a basis for K/L. Consider the set $\{a_i b_j : i \in I, j \in J\}$. We will show that this set is a basis for K/F. If $x \in K$, then $x = \sum_j \alpha_j b_j$ for some $\alpha_j \in L$, with only finitely many of the $b_j \neq 0$. But $\alpha_j = \sum_i \beta_{ij} a_j$ for some $\beta_{ij} \in F$, with only finitely many β_{ij} nonzero for each j. Thus, $x = \sum_{i,j} \beta_{ij} a_i b_j$, so the $\{a_i b_j\}$ span K as an F-vector space. For linear independence, if $\sum_{i,j} \beta_{ij} a_i b_j = 0$ with $\beta_{ij} \in F$, then the independence of the b_j over L shows that $\sum_i \beta_{ij} a_i = 0$ for each j. But independence of the a_i over F gives $\beta_{ij} = 0$ for each i, j. Thus, the $a_i b_j$ are independent over F, so they form a basis for K/F. Therefore,

$$\begin{aligned}[K : F] &= |\{a_i b_j : i \in I, j \in J\}| \\ &= |\{a_i : i \in I\}| \cdot |\{b_j : j \in J\}| = [K : L] \cdot [L : F].\end{aligned}$$

□

This proposition is used primarily with finite extensions, although it is true for arbitrary extensions. Note that the proof above does not assume that the dimensions are finite, although we are being somewhat informal in our treatment of infinite cardinals.

We now prove the converse to Proposition 1.19.

Proposition 1.21 *Let K be a field extension of F. If each $\alpha_i \in K$ is algebraic over F, then $F[\alpha_1, \ldots, \alpha_n]$ is a finite dimensional field extension of F with*

$$[F[\alpha_1, \ldots, \alpha_n] : F] \leq \prod_{i=1}^{n} [F(\alpha_i) : F].$$

Proof. We prove this by induction on n; the case $n = 1$ follows from Proposition 1.15. If we set $L = F[\alpha_1, \ldots, \alpha_{n-1}]$, then by induction L is a field and $[L : F] \leq \prod_{i=1}^{n-1} [F(\alpha_i) : F]$. Then $F[\alpha_1, \ldots, \alpha_n] = L[\alpha_n]$ is a field since α_n is algebraic over L, and since $\min(L, \alpha_n)$ divides $\min(F, \alpha_n)$ by Proposition 1.15, we have $[F[\alpha_1, \ldots, \alpha_n] : L] \leq [F(\alpha_n) : F]$. Hence, by Proposition 1.20 and the induction hypothesis,

$$[F[\alpha_1, \ldots, \alpha_n] : F] = [F[\alpha_1, \ldots, \alpha_n] : L] \cdot [L : F] \leq \prod_{i=1}^{n} [F(\alpha_i) : F].$$

This finishes the proof. □

The inequality of the proposition above can be strict. For example, if $a = \sqrt[4]{2}$ and $b = \sqrt[4]{18}$, then $[\mathbb{Q}(a) : \mathbb{Q}] = [\mathbb{Q}(b) : \mathbb{Q}] = 4$, since the polynomials $x^4 - 2$ and $x^4 - 18$ are irreducible over $\mathbb{Q}$ by an application of the Eisenstein criterion. However, we know that $\mathbb{Q}(a, b) = \mathbb{Q}(\sqrt[4]{2}, \sqrt{3})$, which has degree 8 over $\mathbb{Q}$. To see this equality, note that $(b/a)^4 = 3$, so $(b/a)^2$ is a square root of 3. Thus, $\sqrt{3} \in \mathbb{Q}(a, b)$. However, $[\mathbb{Q}(a, b) : \mathbb{Q}(a)] \leq 2$ because b satisfies the polynomial $x^2 - 3\sqrt{2} = x^2 - 3a^2 \in \mathbb{Q}(a)[x]$. Thus, by Proposition 1.20,

$$[\mathbb{Q}(a, b) : \mathbb{Q}] = [\mathbb{Q}(a, b) : \mathbb{Q}(a)] \cdot [\mathbb{Q}(a) : \mathbb{Q}] \leq 8 = [\mathbb{Q}(\sqrt[4]{2}, \sqrt{3}) : \mathbb{Q}],$$

so since $\mathbb{Q}(\sqrt[4]{2}, \sqrt{3})$ is a subfield of $\mathbb{Q}(a, b)$, we obtain $\mathbb{Q}(a, b) = \mathbb{Q}(\sqrt[4]{2}, \sqrt{3})$. The equality $[\mathbb{Q}(\sqrt[4]{2}, \sqrt{3}) : \mathbb{Q}] = 8$ is left as an exercise (see Problem 18).

As a corollary to the previous proposition, we have the following convenient criterion for an element to be algebraic over a field.

Corollary 1.22 *If K is a field extension of F, then $\alpha \in K$ is algebraic over F if and only if $[F(\alpha) : F] < \infty$. Moreover, K is algebraic over F if $[K : F] < \infty$.*

The converse to the second statement of the corollary is false. There are algebraic extensions of infinite degree. The set of all complex numbers

algebraic over $\mathbb{Q}$ is a field, and this field is infinite dimensional over $\mathbb{Q}$ (see Problem 16).

Proposition 1.21 can be extended easily to the case of fields generated by an arbitrary number of elements.

Proposition 1.23 *Let K be a field extension of F, and let X be a subset of K such that each element of X is algebraic over F. Then $F(X)$ is algebraic over F. If $|X| < \infty$, then $[F(X) : F] < \infty$.*

Proof. Let $a \in F(X)$. By Proposition 1.10, there are $\alpha_1, \dots, \alpha_n \in X$ with $a \in F(\alpha_1, \dots, \alpha_n)$. By Proposition 1.21, $F(\alpha_1, \dots, \alpha_n)$ is algebraic over F. Thus, a is algebraic over F and, hence, $F(X)$ is algebraic over F. If $|X| < \infty$, then $[F(X) : F] < \infty$ by Proposition 1.21. □

We are now ready to prove that the property of being algebraic is transitive. We will use this result frequently. In the case of finite extensions, transitivity follows from Proposition 1.20 and Corollary 1.22, but it is harder to prove for general extensions.

Theorem 1.24 *Let $F \subseteq L \subseteq K$ be fields. If L/F and K/L are algebraic, then K/F is algebraic.*

Proof. Let $\alpha \in K$, and let $f(x) = a_0 + a_1 x + \cdots + x^n$ be the minimal polynomial of α over L. Since L/F is algebraic, the field $L_0 = F(a_0, \dots, a_{n-1})$ is a finite extension of F by Corollary 1.22. Now $f(x) \in L_0[x]$, so α is algebraic over L_0. Thus,

$$[L_0(\alpha) : F] = [L_0(\alpha) : L_0] \cdot [L_0 : F] < \infty.$$

Because $F(\alpha) \subseteq L_0(\alpha)$, we see that $[F(\alpha) : F] < \infty$, so α is algebraic over F. Since this is true for all $\alpha \in K$, we have shown that K/F is algebraic. □

As an application of some of the results we have obtained, we can help to describe the set of algebraic elements of a field extension.

Definition 1.25 *Let K be a field extension of F. The set*

$$\{a \in K : a \text{ is algebraic over } F\}$$

is called the algebraic closure of F in K.

Corollary 1.26 *Let K be a field extension of F, and let L be the algebraic closure of F in K. Then L is a field, and therefore is the largest algebraic extension of F contained in K.*

Proof. Let $a, b \in L$. Then $F(a, b)$ is algebraic over F by Proposition 1.23, so $F(a, b) \subseteq L$, and since $a \pm b, ab, a/b \in F(a, b) \subseteq L$, the set L is closed under the field operations, so it is a subfield of K. Each element of K that is algebraic over F lies in L, which means that L is the largest algebraic extension of F contained in K. □

Composites of field extensions

Let F be a field, and suppose that L_1 and L_2 are field extensions of F contained in some common extension K of F. Then the *composite* L_1L_2 of L_1 and L_2 is the subfield of K generated by L_1 and L_2; that is, $L_1L_2 = L_1(L_2) = L_2(L_1)$. We will use this concept throughout this book. Some properties of composites are given in the Problems. We finish this section with some examples of composites.

Example 1.27 Let $F = \mathbb{Q}$, and view all fields in this example as subfields of $\mathbb{C}$. Let $\omega = e^{2\pi i/3}$, so $\omega^3 = 1$ and $\omega \neq 1$. The composite of $\mathbb{Q}(\sqrt[3]{2})$ and $\mathbb{Q}(\omega\sqrt[3]{2})$ is $\mathbb{Q}(\omega, \sqrt[3]{2})$. To see that this is the composite, note that both $\mathbb{Q}(\sqrt[3]{2})$ and $\mathbb{Q}(\omega\sqrt[3]{2})$ are contained in $\mathbb{Q}(\sqrt[3]{2}, \omega)$, so their composite is also contained in $\mathbb{Q}(\sqrt[3]{2}, \omega)$. However, if a field L contains $\sqrt[3]{2}$ and $\omega\sqrt[3]{2}$, then it also contains $\omega = \omega\sqrt[3]{2}/\sqrt[3]{2}$. Thus, L must contain $\sqrt[3]{2}$ and ω, so it must contain $\mathbb{Q}(\sqrt[3]{2}, \omega)$. Therefore, $\mathbb{Q}(\sqrt[3]{2}, \omega)$ is the smallest field containing both $\mathbb{Q}(\sqrt[3]{2})$ and $\mathbb{Q}(\omega\sqrt[3]{2})$. We can also show that $\mathbb{Q}(\sqrt[3]{2}, \omega) = \mathbb{Q}(\sqrt[3]{2} + \omega)$, so $\mathbb{Q}(\sqrt[3]{2}, \omega)$ is generated by one element over $\mathbb{Q}$. If $a = \omega + \sqrt[3]{2}$, then $(a - \omega)^3 = 2$. Expanding this and using the relation $\omega^2 = -1 - \omega$, solving for ω yields

$$\omega = \frac{a^3 - 3a - 3}{3a^2 + 3a},$$

so $\omega \in \mathbb{Q}(a)$. Thus, $\sqrt{2} = a - \omega \in \mathbb{Q}(a)$, so $\mathbb{Q}(\sqrt[3]{2}, \omega) = \mathbb{Q}(\sqrt[3]{2} + \omega)$.

Example 1.28 The composite of $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt{3})$ is the field $\mathbb{Q}(\sqrt{2}, \sqrt{3})$. This composite can be generated by a single element over $\mathbb{Q}$. In fact, $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\sqrt{2} + \sqrt{3})$. To see this, the inclusion $\supseteq$ is clear. For the reverse inclusion, let $a = \sqrt{2} + \sqrt{3}$. Then $(a - \sqrt{2})^2 = 3$. Multiplying this and rearranging gives $2\sqrt{2}a = a^2 - 1$, so

$$\sqrt{2} = \frac{a^2 - 1}{2a} \in \mathbb{Q}(a).$$

Similar calculations show that

$$\sqrt{3} = \frac{(a^2 + 1)}{2a} \in \mathbb{Q}(a).$$

Therefore, $\mathbb{Q}(\sqrt{2}, \sqrt{3}) \subseteq \mathbb{Q}(a)$, which, together with the previous inclusion, gives $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(a)$.

We will see in Section 5 that every finite extension of $\mathbb{Q}$ is of the form $\mathbb{Q}(a)$ for some a, which indicates that there is some reason behind these ad hoc calculations.

Problems

1. Let K be a field extension of F. By defining scalar multiplication for $\alpha \in F$ and $a \in K$ by $\alpha \cdot a = \alpha a$, the multiplication in K, show that K is an F-vector space.

2. If K is a field extension of F, prove that $[K : F] = 1$ if and only if $K = F$.

3. Let K be a field extension of F, and let $a \in K$. Show that the evaluation map $\text{ev}_a : F[x] \to K$ given by $\text{ev}_a(f(x)) = f(a)$ is a ring and an F-vector space homomorphism.
   (Such a map is called an *F-algebra homomorphism.*)

4. Prove Proposition 1.9.

5. Show that $\mathbb{Q}(\sqrt{5}, \sqrt{7}) = \mathbb{Q}(\sqrt{5} + \sqrt{7})$.

6. Verify the following universal mapping property for polynomial rings:

   (a) Let A be a ring containing a field F. If $a_1, \dots, a_n \in A$, show that there is a unique ring homomorphism $\varphi : F[x_1, \dots, x_n] \to A$ with $\varphi(x_i) = a_i$ for each i.

   (b) Moreover, suppose that B is a ring containing F, together with a function $f : \{x_1, \dots, x_n\} \to B$, satisfying the following property: For any ring A containing F and elements $a_1, \dots, a_n \in A$, there is a unique ring homomorphism $\varphi : B \to A$ with $\varphi(f(x_i)) = a_i$. Show that B is isomorphic to $F[x_1, \dots, x_n]$.

7. Let A be a ring. If A is also an F-vector space and $\alpha(ab) = (\alpha a)b = a(\alpha b)$ for all $\alpha \in F$ and $a, b \in A$, then A is said to be an *F-algebra.* If A is an F-algebra, show that A contains an isomorphic copy of F. Also show that if K is a field extension of F, then K is an F-algebra.

8. Let $K = F(a)$ be a finite extension of F. For $\alpha \in K$, let L_α be the map from K to K defined by $L_\alpha(x) = \alpha x$. Show that L_α is an F-linear transformation. Also show that $\det(xI - L_a)$ is the minimal polynomial $\min(F, a)$ of a. For which $\alpha \in K$ is $\det(xI - L_\alpha) = \min(F, \alpha)$?

9. If K is an extension of F such that $[K : F]$ is prime, show that there are no intermediate fields between K and F.

10. If K is a field extension of F and if $a \in K$ such that $[F(a) : F]$ is odd, show that $F(a) = F(a^2)$. Give an example to show that this can be false if the degree of $F(a)$ over F is even.

11. If K is an algebraic extension of F and if R is a subring of K with $F \subseteq R \subseteq K$, show that R is a field.

12. Show that $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt{3})$ are not isomorphic as fields but are isomorphic as vector spaces over $\mathbb{Q}$.

13. If $L_1 = F(a_1, \ldots, a_n)$ and $L_2 = F(b_1, \ldots, b_m)$, show that the composite L_1L_2 is equal to $F(a_1, \ldots, a_n, b_1, \ldots, b_m)$.

14. If L_1 and L_2 are field extensions of F that are contained in a common field, show that L_1L_2 is a finite extension of F if and only if both L_1 and L_2 are finite extensions of F.

15. If L_1 and L_2 are field extensions of F that are contained in a common field, show that L_1L_2 is algebraic over F if and only if both L_1 and L_2 are algebraic over F.

16. Let $\mathbb{A}$ be the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$. Prove that $[\mathbb{A} : \mathbb{Q}] = \infty$.

17. Let K be a finite extension of F. If L_1 and L_2 are subfields of K containing F, show that $[L_1L_2 : F] \le [L_1 : F] \cdot [L_2 : F]$. If $\gcd([L_1 : F], [L_2 : F]) = 1$, prove that $[L_1L_2 : F] = [L_1 : F] \cdot [L_2 : F]$.

18. Show that $[\mathbb{Q}(\sqrt[4]{2}, \sqrt{3}) : \mathbb{Q}] = 8$.

19. Give an example of field extensions L_1, L_2 of F for which $[L_1L_2 : F] < [L_1 : F] \cdot [L_2 : F]$.

20. Give an example of a field extension K/F with $[K : F] = 3$ but with $K \neq F(\sqrt[3]{b})$ for any $b \in F$.

21. Let $a \in \mathbb{C}$ be a root of $x^n - b$, where $b \in \mathbb{C}$. Show that $x^n - b$ factors as $\prod_{i=0}^{n-1}(x - \omega^i a)$, where $\omega = e^{2\pi i/n}$.

22. (a) Let F be a field, and let $f(x) \in F[x]$. If $f(x) = \sum_i a_i x^i$ and $\alpha \in F$, let $f(x+\alpha) = \sum_i a_i (x+\alpha)^i$. Prove that f is irreducible over F if and only if $f(x+\alpha)$ is irreducible over F for any $\alpha \in F$.

    (b) Show that $x^{p-1} + x^{p-2} + \cdots + x + 1$ is irreducible over $\mathbb{Q}$ if p is a prime.
    (Hint: Replace x by $x + 1$ and use the Eisenstein criterion.)

23. Recall that the *characteristic* of a ring R with identity is the smallest positive integer n for which $n \cdot 1 = 0$, if such an n exists, or else the characteristic is 0. Let R be a ring with identity. Define $\varphi : \mathbb{Z} \to R$ by $\varphi(n) = n \cdot 1$, where 1 is the identity of R. Show that φ is a

ring homomorphism and that $\ker(\varphi) = m\mathbb{Z}$ for a unique nonnegative integer m, and show that m is the characteristic of R.

24. For any positive integer n, give an example of a ring of characteristic n.

25. If R is an integral domain, show that either $\text{char}(R) = 0$ or $\text{char}(R)$ is prime.

26. Let R be a commutative ring with identity. The *prime subring of* R is the intersection of all subrings of R. Show that this intersection is a subring of R that is contained inside all subrings of R. Moreover, show that the prime subring of R is equal to $\{n \cdot 1 : n \in \mathbb{Z}\}$, where 1 is the multiplicative identity of R.

27. Let F be a field. If $\text{char}(F) = p > 0$, show that the prime subring of R is isomorphic to the field $\mathbb{F}_p$, and if $\text{char}(F) = 0$, then the prime subring is isomorphic to $\mathbb{Z}$.

28. Let F be a field. The *prime subfield of* F is the intersection of all subfields of F. Show that this subfield is the quotient field of the prime subring of F, that it is contained inside all subfields of F, and that it is isomorphic to $\mathbb{F}_p$ or $\mathbb{Q}$ depending on whether the characteristic of F is $p > 0$ or 0.

2 Automorphisms

The main idea of Galois was to associate to any polynomial f a group of permutations of the roots of f. In this section, we define and study this group and give some numerical information about it. Our description of this group is not the one originally given by Galois but an equivalent description given by Artin.

Let K be a field. A ring isomorphism from K to K is usually called an *automorphism* of K. The group of all automorphisms of K will be denoted $\text{Aut}(K)$. Because we are interested in field extensions, we need to consider mappings of extensions. Let K and L be extension fields of F. An *F-homomorphism* $\tau : K \longrightarrow L$ is a ring homomorphism such that $\tau(a) = a$ for all $a \in F$; that is, $\tau|_F = \text{id}$. If τ is a bijection, then τ is called an *F-isomorphism.* An F-isomorphism from a field K to itself is called an *F-automorphism.*

Let us point out some simple properties of F-homomorphisms. If $\tau : K \to L$ is an F-homomorphism of extension fields of F, then τ is also a linear transformation of F-vector spaces, since $\tau(\alpha a) = \tau(\alpha)\tau(a) = \alpha\tau(a)$ for $\alpha \in F$ and $a \in K$. Furthermore, $\tau \neq 0$, so τ is injective since K is a field. Also, if $[K : F] = [L : F] < \infty$, then τ is automatically surjective by

dimension counting. In particular, any F-homomorphism from K to itself is a bijection, provided that $[K : F] < \infty$.

Definition 2.1 *Let K be a field extension of F. The Galois group* $\mathrm{Gal}(K/F)$ *is the set of all F-automorphisms of K.*

If $K = F(X)$ is generated over F by a subset X, we can determine the F-automorphisms of K in terms of their action on the generating set X. For instance, if K is an extension of F that is generated by the roots of a polynomial $f(x) \in F[x]$, the following two lemmas will allow us to interpret the Galois group $\mathrm{Gal}(K/F)$ as a group of permutations of the roots of f. This type of field extension obtained by adjoining to a base field roots of a polynomial is extremely important, and we will study it in Section 3. One use of these two lemmas will be to help calculate Galois groups, as shown in the examples below.

Lemma 2.2 *Let $K = F(X)$ be a field extension of F that is generated by a subset X of K. If $\sigma, \tau \in \mathrm{Gal}(K/F)$ with $\sigma|_X = \tau|_X$, then $\sigma = \tau$. Therefore, F-automorphisms of K are determined by their action on a generating set.*

Proof. Let $a \in K$. Then there is a finite subset $\{\alpha_1, \ldots, \alpha_n\} \subseteq X$ with $a \in F(\alpha_1, \ldots, \alpha_n)$. This means there are polynomials $f, g \in F[x_1, \ldots, x_n]$ with $a = f(\alpha_1, \ldots, \alpha_n)/g(\alpha_1, \ldots, \alpha_n)$; say

$$f(x_1, \ldots, x_n) = \sum b_{i_1 i_2 \cdots i_n} x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n},$$
$$g(x_1, \ldots, x_n) = \sum c_{i_1 i_2 \cdots i_n} x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n},$$

where each coefficient is in F. Since σ and τ preserve addition and multiplication, and fix elements of F, we have

$$\begin{aligned}\sigma(a) &= \sum \frac{b_{i_1 i_2 \cdots i_n} \sigma(\alpha_1)^{i_1} \sigma(\alpha_2)^{i_2} \cdots \sigma(\alpha_n)^{i_n}}{c_{i_1 i_2 \cdots i_n} \sigma(\alpha_1)^{i_1} \sigma(\alpha_2)^{i_2} \cdots \sigma(\alpha_n)^{i_n}} \\ &= \sum \frac{b_{i_1 i_2 \cdots i_n} \tau(\alpha_1)^{i_1} \tau(\alpha_2)^{i_2} \cdots \tau(\alpha_n)^{i_n}}{c_{i_1 i_2 \cdots i_n} \tau(\alpha_1)^{i_1} \tau(\alpha_2)^{i_2} \cdots \tau(\alpha_n)^{i_n}} \\ &= \tau(a).\end{aligned}$$

Thus, $\sigma = \tau$, so F-automorphisms are determined by their action on generators. □

Lemma 2.3 *Let $\tau : K \longrightarrow L$ be an F-homomorphism and let $\alpha \in K$ be algebraic over F. If $f(x)$ is a polynomial over F with $f(\alpha) = 0$, then $f(\tau(\alpha)) = 0$. Therefore, τ permutes the roots of* $\min(F, \alpha)$. *Also,* $\min(F, \alpha) = \min(F, \tau(\alpha))$.

Proof. Let $f(x) = a_0 + a_1 x + \cdots + a_n x^n$. Then

$$0 = \tau(0) = \tau(f(\alpha)) = \sum_i \tau(a_i)\tau(\alpha)^i.$$

But, since each $a_i \in F$, we have $\tau(a_i) = a_i$. Thus, $0 = \sum_i a_i \tau(\alpha)^i$, so $f(\tau(\alpha)) = 0$. In particular, if $p(x) = \min(F, \alpha)$, then $p(\tau(\alpha)) = 0$, so $\min(F, \tau(\alpha))$ divides $p(x)$. Since $p(x)$ is irreducible, $\min(F, \tau(\alpha)) = p(x) = \min(F, \alpha)$. □

Corollary 2.4 *If* $[K : F] < \infty$, *then* $|\operatorname{Gal}(K/F)| < \infty$.

Proof. We can write $K = F(\alpha_1, \ldots, \alpha_n)$ for some $\alpha_i \in K$. Any F-automorphism of K is determined by what it does to the α_i. By Lemma 2.3, there are only finitely many possibilities for the image of any α_i; hence, there are only finitely many automorphisms of K/F. □

Example 2.5 Consider the extension $\mathbb{C}/\mathbb{R}$. We claim that $\operatorname{Gal}(\mathbb{C}/\mathbb{R}) = \{\mathrm{id}, \sigma\}$, where σ is complex conjugation. Both of these functions are $\mathbb{R}$-automorphisms of $\mathbb{C}$, so they are contained in $\operatorname{Gal}(\mathbb{C}/\mathbb{R})$. To see that there is no other automorphism of $\mathbb{C}/\mathbb{R}$, note that an element of $\operatorname{Gal}(\mathbb{C}/\mathbb{R})$ is determined by its action on i, since $\mathbb{C} = \mathbb{R}(i)$. Lemma 2.3 shows that if $\tau \in \operatorname{Gal}(\mathbb{C}/\mathbb{R})$, then $\tau(i)$ is a root of $x^2 + 1$, so $\tau(i)$ must be either i or $-i$. Therefore, $\tau = \mathrm{id}$ or $\tau = \sigma$.

Example 2.6 The Galois group of $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ is $\langle \mathrm{id} \rangle$. To see this, if σ is a $\mathbb{Q}$-automorphism of $\mathbb{Q}(\sqrt[3]{2})$, then $\sigma(\sqrt[3]{2})$ is a root of $\min(\mathbb{Q}, \sqrt[3]{2}) = x^3 - 2$. If $\omega = e^{2\pi i/3}$, then the roots of this polynomial are $\sqrt[3]{2}$, $\omega\sqrt[3]{2}$, and $\omega^2\sqrt[3]{2}$. The only root of $x^3 - 2$ that lies in $\mathbb{Q}(\sqrt[3]{2})$ is $\sqrt[3]{2}$, since if another root lies in this field, then $\omega \in \mathbb{Q}(\sqrt[3]{2})$, which is false since $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$ and $[\mathbb{Q}(\omega) : \mathbb{Q}] = 2$. Therefore, $\sigma(\sqrt[3]{2}) = \sqrt[3]{2}$, and since σ is determined by its action on the generator $\sqrt[3]{2}$, we see that $\sigma = \mathrm{id}$.

Example 2.7 Let $K = \mathbb{F}_2(t)$ be the rational function field in one variable over $\mathbb{F}_2$, and let $F = \mathbb{F}_2(t^2)$. Then $[K : F] = 2$. The element t satisfies the polynomial $x^2 - t^2 \in F[x]$, which has only t as a root, since $x^2 - t^2 = (x - t)^2$ in $K[x]$. Consequently, if σ is an F-automorphism of K, then $\sigma(t) = t$, so $\sigma = \mathrm{id}$. This proves that $\operatorname{Gal}(K/F) = \{\mathrm{id}\}$.

Example 2.8 Let $F = \mathbb{F}_2$. The polynomial $1 + x + x^2$ is irreducible over F, since it has no roots in F. In fact, this is the only irreducible quadratic over F; the three other quadratics factor over F. Let $K = F[x]/(1 + x + x^2)$, a field that we can view as an extension field of F; see Example 1.6 for details on this construction. To simplify notation, we write $M = (1 + x + x^2)$. Every element of K can be written in the form $a + bx + M$ by the division algorithm. Let us write $\alpha = x + M$. The subfield $\{a + M : a \in F\}$ of K is

isomorphic to F. By identifying F with this subfield of K, we can write every element of K in the form $a+b\alpha$ with $a, b \in F$. Then $K = F(\alpha)$, so any F-automorphism of K is determined by its action on α. By Lemma 2.3, if σ is an F-automorphism of K, then $\sigma(\alpha)$ is a root of $1+x+x^2$. By factoring $1+x+x^2$ as $(x-\alpha)(x-\beta)$ and expanding, we see that the other root of $1+x+x^2$ is $\alpha+1$. Therefore, the only possibility for $\sigma(\alpha)$ is α or $\alpha+1$, so $\mathrm{Gal}(K/F)$ has at most two elements. To see that $\mathrm{Gal}(K/F)$ has exactly two elements, we need to check that there is indeed an automorphism σ with $\sigma(\alpha) = \alpha+1$. If σ does exist, then $\sigma(a+b\alpha) = a+b(\alpha+1) = (a+b)+b\alpha$. We leave it as an exercise (Problem 7) to show that the function $\sigma : K \to K$ defined by $\sigma(a+b\alpha) = (a+b)+b\alpha$ is an F-automorphism of K. Therefore, $\mathrm{Gal}(K/F) = \{\mathrm{id}, \sigma\}$.

The idea of Galois theory is to be able to go back and forth from field extensions to groups. We have now seen how to take a field extension K/F and associate a group, $\mathrm{Gal}(K/F)$. More generally, if L is a field with $F \subseteq L \subseteq K$, we can associate a group $\mathrm{Gal}(K/L)$. This is a subgroup of $\mathrm{Gal}(K/F)$, as we will see in the lemma below. Conversely, given a subgroup of $\mathrm{Gal}(K/F)$ we can associate a subfield of K containing F. Actually, we can do this for an arbitrary subset of $\mathrm{Aut}(K)$. Let S be a subset of $\mathrm{Aut}(K)$, and set

$$\mathcal{F}(S) = \{\, a \in K \,:\, \tau(a) = a \text{ for all } \tau \in S\} \ .$$

It is not hard to see that $\mathcal{F}(S)$ is a subfield of K, called the *fixed field* of S. A field L with $F \subseteq L \subseteq K$ is called an *intermediate field* of the extension K/F. Therefore, if $S \subseteq \mathrm{Gal}(K/F)$, then $\mathcal{F}(S)$ is an intermediate field of K/F.

The following lemma gives some simple properties of Galois groups and fixed fields.

Lemma 2.9 *Let K be a field.*

1. *If $L_1 \subseteq L_2$ are subfields of K, then* $\mathrm{Gal}(K/L_2) \subseteq \mathrm{Gal}(K/L_1)$.
2. *If L is a subfield of K, then* $L \subseteq \mathcal{F}(\mathrm{Gal}(K/L))$.
3. *If $S_1 \subseteq S_2$ are subsets of* $\mathrm{Aut}(K)$, *then* $\mathcal{F}(S_2) \subseteq \mathcal{F}(S_1)$.
4. *If S is a subset of* $\mathrm{Aut}(K)$, *then* $S \subseteq \mathrm{Gal}(K/\mathcal{F}(S))$.
5. *If $L = \mathcal{F}(S)$ for some $S \subseteq$* $\mathrm{Aut}(K)$, *then* $L = \mathcal{F}(\mathrm{Gal}(K/L))$.
6. *If* $H = \mathrm{Gal}(K/L)$ *for some subfield L of K, then* $H = \mathrm{Gal}(K/\mathcal{F}(H))$.

Proof. The first four parts are simple consequences of the definitions. We leave the proofs of parts 2, 3, and 4 to the reader and prove part 1 for the sake of illustration. If $\sigma \in \mathrm{Gal}(K/L_2)$, then $\sigma(a) = a$ for all $a \in L_2$. Thus, $\sigma(a) = a$ for all $a \in L_1$, as $L_1 \subseteq L_2$, so $\sigma \in \mathrm{Gal}(K/L_1)$.

To prove part 5, suppose that $L = \mathcal{F}(S)$ for some subset S of $\text{Aut}(K)$. Then $S \subseteq \text{Gal}(K/L)$, so $\mathcal{F}(\text{Gal}(K/L)) \subseteq \mathcal{F}(S) = L$. But $L \subseteq \mathcal{F}(\text{Gal}(K/L))$, so $L = \mathcal{F}(\text{Gal}(K/L))$. For part 6, if $H = \text{Gal}(K/L)$ for some subfield L of K, then $L \subseteq \mathcal{F}(\text{Gal}(K/L))$, so

$$\text{Gal}(K/\mathcal{F}(\text{Gal}(K/L))) \subseteq \text{Gal}(K/L) = H.$$

However, by part 4 we have $H \subseteq \text{Gal}(K/\mathcal{F}(H))$, so $H = \text{Gal}(K/\mathcal{F}(H))$. □

Corollary 2.10 *If K is a field extension of F, then there is* 1–1 *inclusion reversing correspondence between the set of subgroups of* $\text{Gal}(K/F)$ *of the form* $\text{Gal}(K/L)$ *for some subfield L of K containing F and the set of subfields of K that contain F of the form $\mathcal{F}(S)$ for some subset S of* $\text{Aut}(K)$*. This correspondence is given by $L \mapsto \text{Gal}(K/L)$, and its inverse is given by $H \mapsto \mathcal{F}(H)$.*

Proof. This follows immediately from the lemma. If $\mathcal{G}$ and $\mathcal{F}$ are respectively the set of groups and fields in question, then the map that sends a subfield L of K to the subgroup $\text{Gal}(K/L)$ of $\text{Aut}(K)$ sends $\mathcal{F}$ to $\mathcal{G}$. This map is injective and surjective by part 5 of the lemma. Its inverse is given by sending H to $\mathcal{F}(H)$ by part 6. □

If K/F is a finite extension, under what circumstances does the association $L \mapsto \text{Gal}(K/L)$ give an inclusion reversing correspondence between the set of all subfields of K containing F and the set of all subgroups of $\text{Gal}(K/F)$? A necessary condition from part 5 is that $F = \mathcal{F}(\text{Gal}(K/F))$. We shall see in Section 5 that this is actually a sufficient condition.

The next three results aim at getting more precise numerical information on $|\text{Gal}(K/F)|$ for a finite extension K/F. We first need a definition.

Definition 2.11 *If G is a group and if K is a field, then a character is a group homomorphism from G to K^*.*

By setting $G = K^*$, we see that F-automorphisms of K can be viewed as characters from G to K^*. The next lemma will lead to a bound on $|\text{Gal}(K/F)|$.

Lemma 2.12 (Dedekind's Lemma) *Let $\tau_1, \dots, \tau_n$ be distinct characters from G to K^*. Then the τ_i are linearly independent over K; that is, if $\sum_i c_i \tau_i(g) = 0$ for all $g \in G$, where the $c_i \in K$, then all $c_i = 0$.*

Proof. Suppose that the lemma is false. Choose k minimal (relabeling the τ_i if necessary) so that there are $c_i \in K$ with $\sum_i c_i \tau_i(g) = 0$ for all $g \in G$. Then all $c_i \neq 0$. Since $\tau_1 \neq \tau_2$, there is an $h \in G$ with $\tau_1(h) \neq \tau_2(h)$. We

have $\sum_{i=1}^{k}(c_i\tau_1(h))\tau_i(g) = 0$ and

$$\sum_{i=1}^{k} c_i\tau_i(hg) = \sum_i (c_i\tau_i(h))\tau_i(g) = 0$$

for all g. Subtracting gives $\sum_{i=1}^{k}(c_i(\tau_1(h) - \tau_i(h)))\tau_i(g) = 0$ for all g. This is an expression involving $k-1$ of the τ_i with not all of the coefficients zero. This contradicts the minimality of k, so the lemma is proved. □

There is a vector space interpretation of Dedekind's lemma. If V is the set of all functions from G to K, then V is a K-vector space under usual function addition and scalar multiplication, and Dedekind's lemma can be viewed as showing that the set of characters from G to K^* forms a linearly independent set in V.

Proposition 2.13 *If K is a finite field extension of F, then* $|\operatorname{Gal}(K/F)| \le [K : F]$.

Proof. The group $\operatorname{Gal}(K/F)$ is finite by Corollary 2.4. Let $\operatorname{Gal}(K/F) = \{\tau_1, \ldots, \tau_n\}$, and suppose that $[K : F] < n$. Let $\alpha_1, \ldots, \alpha_m$ be a basis for K as an F-vector space. The matrix

$$A = \begin{pmatrix} \tau_1(\alpha_1) & \tau_1(\alpha_2) & \cdots & \tau_1(\alpha_m) \\ \tau_2(\alpha_1) & \tau_2(\alpha_2) & \cdots & \tau_2(\alpha_m) \\ \vdots & \vdots & \ddots & \vdots \\ \tau_n(\alpha_1) & \tau_n(\alpha_2) & \cdots & \tau_n(\alpha_m) \end{pmatrix}$$

over K has $\operatorname{rank}(A) \le m < n$, so the rows of A are linearly dependent over K. Thus, there are $c_i \in K$, not all zero, such that $\sum_i c_i\tau_i(\alpha_j) = 0$ for all j. If we set $G = K^*$, then for $g \in G$ there are $a_i \in F$ with $g = \sum_j a_j\alpha_j$. Thus,

$$\begin{aligned} \sum_i c_i\tau_i(g) &= \sum_i c_i\tau_i\left(\sum_j a_j\alpha_j\right) = \sum_i c_i\left(a_j\sum_j \tau_j(\alpha_j)\right) \\ &= \sum_j a_j\left(\sum_i c_i\tau_i(\alpha_j)\right) = 0. \end{aligned}$$

All the c_i are then 0 by Dedekind's lemma. This contradiction proves that $\operatorname{Gal}(K/F) \le [K : F]$. □

The following question arises naturally from this proposition: For which field extensions K/F does $|\operatorname{Gal}(K/F)| = [K : F]$? The inequality in the proposition above may be strict, as shown in Examples 2.6 and 2.7.

The next proposition determines when $|\operatorname{Gal}(K/F)| = [K : F]$, provided that the group $\operatorname{Gal}(K/F)$ is finite.

Proposition 2.14 *Let G be a finite group of automorphisms of K with $F = \mathcal{F}(G)$. Then $|G| = [K : F]$, and so $G = \mathrm{Gal}(K/F)$.*

Proof. By the previous proposition, $|G| \leq [K : F]$ since $G \subseteq \mathrm{Gal}(K/F)$. Suppose that $|G| < [K : F]$. Let $n = |G|$, and take $\alpha_1, \ldots, \alpha_{n+1} \in K$ linearly independent over F. If $G = \{\tau_1, \ldots, \tau_n\}$, let A be the matrix

$$A = \begin{pmatrix} \tau_1(\alpha_1) & \tau_1(\alpha_2) & \cdots & \tau_1(\alpha_{n+1}) \\ \tau_2(\alpha_1) & \tau_2(\alpha_2) & \cdots & \tau_2(\alpha_{n+1}) \\ \vdots & \vdots & \ddots & \vdots \\ \tau_n(\alpha_1) & \tau_n(\alpha_2) & \cdots & \tau_n(\alpha_{n+1}) \end{pmatrix}.$$

Then the columns of A are linearly dependent over K. Choose k minimal so that the first k columns of A are linearly dependent over K (relabeling if necessary). Thus, there are $c_i \in K$ not all zero with $\sum_{i=1}^k c_i \tau_j(\alpha_i) = 0$ for all j. Minimality of k shows all $c_i \neq 0$. Thus, by dividing we may assume that $c_1 = 1$. If each $c_i \in F$, then $0 = \tau_j(\sum_{i=1}^k c_i \alpha_i)$ for each j, so $\sum_{i=1}^k c_i \alpha_i = 0$. This is false by the independence of the α_i over F. Take $\sigma \in G$. Since σ permutes the elements of G, we get $\sum_{i=1}^k \sigma(c_i)\tau_j(\alpha_i) = 0$ for all j. Subtracting this from the original equation and recalling that $c_1 = 1$ gives $\sum_{i=2}^k (c_i - \sigma(c_i))\tau_j(\alpha_i) = 0$ for all j. Minimality of k shows that $c_i - \sigma(c_i) = 0$ for each i. Since this is true for all $\sigma \in G$, we get all $c_i \in \mathcal{F}(G) = F$. But we have seen that this leads to a contradiction. Thus $|G| = [K : F]$. In particular, $G = \mathrm{Gal}(K/F)$, since $G \subseteq \mathrm{Gal}(K/F)$ and $|G| = [K : F] \geq |\,\mathrm{Gal}(K/F)|$. □

The field extensions described in Proposition 2.14 are those of particular interest to us, as they were to Galois in his work on the solvability of polynomials.

Definition 2.15 *Let K be an algebraic extension of F. Then K is Galois over F if $F = \mathcal{F}(\mathrm{Gal}(K/F))$.*

If $[K : F] < \infty$, then Proposition 2.14 gives us a numerical criterion for when K/F is Galois.

Corollary 2.16 *Let K be a finite extension of F. Then K/F is Galois if and only if $|\mathrm{Gal}(K/F)| = [K : F]$.*

Proof. If K/F is a Galois extension, then $F = \mathcal{F}(\mathrm{Gal}(K/F))$, so by Proposition 2.14, $|\mathrm{Gal}(K/F)| = [K : F]$. Conversely, if $|\mathrm{Gal}(K/F)| = [K : F]$, let $L = \mathcal{F}(\mathrm{Gal}(K/F))$. Then $\mathrm{Gal}(K/L) = \mathrm{Gal}(K/F)$ by Proposition 2.14, and so $|\mathrm{Gal}(K/F)| = [K : L] \leq [K : F]$. Since $|\mathrm{Gal}(K/F)| = [K : F]$, this forces $[K : L] = [K : F]$, so $L = F$. □

The previous corollary gives us a numerical criterion for when a finite extension is Galois. However, to use it we need to know the Galois group of the extension. This group is not always easy to determine. For extensions of F of the form $F(a)$, we have a simpler criterion to determine when $F(a)/F$ is Galois.

Corollary 2.17 *Let K be a field extension of F, and let $a \in K$ be algebraic over F. Then $|\mathrm{Gal}(F(a)/F)|$ is equal to the number of distinct roots of $\min(F,a)$ in $F(a)$. Therefore, $F(a)$ is Galois over F if and only if $\min(F,a)$ has n distinct roots in $F(a)$, where $n = \deg(\min(F,a))$.*

Proof. If $\tau \in \mathrm{Gal}(F(a)/F)$, we have seen that $\tau(a)$ is a root of $\min(F,a)$. Moreover, if $\sigma, \tau \in \mathrm{Gal}(F(a)/F)$ with $\sigma \neq \tau$, then $\sigma(a) \neq \tau(a)$, since F-automorphisms on $F(a)$ are determined by their action on a. Therefore, $|\mathrm{Gal}(F(a)/F)| \leq n$. Conversely, let b be a root in $F(a)$ of $\min(F,a)$. Define $\tau : F(a) \to F(a)$ by $\tau(f(a)) = f(b)$ for any $f(x) \in F[x]$. This map is well defined precisely because b is a root of $\min(F,a)$. It is straightforward to show that τ is an F-automorphism, and $\tau(a) = b$ by the definition of τ. Thus, $|\mathrm{Gal}(F(a)/F)|$ is equal to the number of distinct roots of $\min(F,a)$ in $F(a)$. Since $[F(a) : F] = \deg(\min(F,a))$, we see that $F(a)$ is Galois over F if and only if $\min(F,a)$ has n distinct roots in $F(a)$. □

There are two ways that a field extension $F(a)/F$ can fail to be Galois. First, if $p(x) = \min(F,a)$, then p could fail to have all its roots in $F(a)$. Second, $p(x)$ could have repeated roots. The next two sections will address these concerns. We finish this section with a number of examples of extensions for which we determine whether or not they are Galois. Here and elsewhere in this book, we use the idea of the characteristic of a field (or a ring with identity). For the reader unfamiliar with this notion, the characteristic $\mathrm{char}(F)$ of a field F is the order of the multiplicative identity 1 as an element of the additive group $(F,+)$, provided that this order is finite, or else $\mathrm{char}(F) = 0$ if this order is infinite. Note that the characteristic of a field is either 0 or is a prime number. More information on the characteristic of a ring can be found in Appendix A or in the last six problems in the previous section.

Example 2.18 The extension $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ is not Galois, for we have seen that $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$ but $\left|\mathrm{Gal}(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})\right| = 1$. The polynomial $x^3 - 2$ has three distinct roots, but only one of them lies in $\mathbb{Q}(\sqrt[3]{2})$.

Example 2.19 Let k be a field of characteristic $p > 0$, and let $k(t)$ be the rational function field in one variable over k. Consider the field extension $k(t)/k(t^p)$. Then t satisfies the polynomial $x^p - t^p \in k(t^p)[x]$. However, over $k(t)$ this polynomial factors as $x^p - t^p = (x-t)^p$. Thus, the minimal polynomial of t over $k(t^p)$ has only one root; consequently, $\mathrm{Gal}(k(t)/k(t^p)) = \{\mathrm{id}\}$. Thus, $k(t)/k(t^p)$ is not Galois.

The previous two examples illustrate the two ways a field extension of the form $F(a)/F$ can fail to be Galois. The remaining examples are examples of extensions that are Galois.

Example 2.20 Let F be a field of characteristic not 2, and let $a \in F$ be an element that is not the square of any element in F. Let $K = F[x]/(x^2 - a)$, a field since $x^2 - a$ is irreducible over F. We view F as a subfield of K by identifying F with the subfield $\{\alpha + (x^2 - a) : \alpha \in F\}$ of K. Under this identification, each coset is uniquely expressible in the form $\alpha + \beta x + (x^2 - a)$ and, hence, is an F-linear combination of $1 + (x^2 - a)$ and $x + (x^2 - a)$. Thus, 1 and $u = x + (x^2 - a)$ form a basis for K as an F-vector space, so $[K : F] = 2$. If σ is defined by

$$\sigma(\alpha + \beta u) = \alpha - \beta u,$$

then σ is an automorphism of K since u and $-u$ are roots in K of $x^2 - a$. Thus, $\mathrm{id}, \sigma \in \mathrm{Gal}(K/F)$, so $|\mathrm{Gal}(K/F)| = 2 = [K : F]$. Consequently, K/F is a Galois extension.

The extension $K = F(\alpha)$ is generated by an element α with $\alpha^2 = a$. We will often write $F(\sqrt{a})$ for this extension. The notation $\sqrt{a}$ is somewhat ambiguous, since for an arbitrary field F there is no way to distinguish between different square roots, although this will not cause us any problems.

Example 2.21 The extension $\mathbb{Q}(\sqrt[3]{2}, \omega)/\mathbb{Q}$ is Galois, where $\omega = e^{2\pi i/3}$. In fact, the field $\mathbb{Q}(\sqrt[3]{2}, \omega)$ is the field generated over $\mathbb{Q}$ by the three roots $\sqrt[3]{2}$, $\omega\sqrt[3]{2}$, and $\omega^2\sqrt[3]{2}$, of $x^3 - 2$, and since ω satisfies $x^2 + x + 1$ over $\mathbb{Q}$ and ω is not in $\mathbb{Q}(\sqrt[3]{2})$, we see that $[\mathbb{Q}(\sqrt[3]{2}, \omega) : \mathbb{Q}] = 6$. It can be shown (see Problem 3) that the six functions

$$\begin{aligned}
\mathrm{id} &: \sqrt[3]{2} \to \sqrt[3]{2}, \quad \omega \to \omega,\\
\sigma &: \sqrt[3]{2} \to \omega\sqrt[3]{2}, \quad \omega \to \omega,\\
\tau &: \sqrt[3]{2} \to \sqrt[3]{2}, \quad \omega \to \omega^2,\\
\rho &: \sqrt[3]{2} \to \omega\sqrt[3]{2}, \quad \omega \to \omega^2,\\
\mu &: \sqrt[3]{2} \to \omega^2\sqrt[3]{2}, \quad \omega \to \omega,\\
\xi &: \sqrt[3]{2} \to \omega^2\sqrt[3]{2}, \quad \omega \to \omega^2
\end{aligned}$$

extend to distinct automorphisms of $\mathbb{Q}(\sqrt[3]{2}, \omega)/\mathbb{Q}$. Thus,

$$\left|\mathrm{Gal}(\mathbb{Q}(\sqrt[3]{2}, \omega)/\mathbb{Q})\right| = [\mathbb{Q}(\sqrt[3]{2}, \omega) : \mathbb{Q}],$$

and so $\mathbb{Q}(\omega, \sqrt[3]{2})/\mathbb{Q}$ is Galois.

One reason we did not do the calculation that shows that we do get six automorphisms from these formulas is that this calculation is long and not particularly informative. Another reason is that later on we will see

easier ways to determine when an extension is Galois. Knowing ahead of time that $\mathbb{Q}(\sqrt[3]{2}, \omega)/\mathbb{Q}$ is Galois and that the degree of this extension is six tells us that we have six $\mathbb{Q}$-automorphisms of $\mathbb{Q}(\sqrt[3]{2}, \omega)$. There are only six possibilities for the images of $\sqrt[3]{2}$ and ω under an automorphism, and so all six must occur.

Example 2.22 This example shows us that any finite group can occur as the Galois group of a Galois extension. We will use this example a number of times in later sections. Let k be a field and let $K = k(x_1, x_2, \ldots, x_n)$ be the field of rational functions in n variables over k. For each permutation $\sigma \in S_n$, define $\sigma(x_i) = x_{\sigma(i)}$. Then σ has a natural extension to an automorphism of K by defining

$$\sigma\left(\frac{f(x_1, \ldots, x_n)}{g(x_1, \ldots, x_n)}\right) = \frac{f(x_{\sigma(1)}, \ldots, x_{\sigma(n)})}{g(x_{\sigma(1)}, \ldots, x_{\sigma(n)})}.$$

The straightforward but somewhat messy calculation that this does define a field automorphism on K is left to Problem 5. We can then view $S_n \subseteq \text{Aut}(K)$. Let $F = \mathcal{F}(S_n)$. By Proposition 2.14, K/F is a Galois extension with $\text{Gal}(K/F) = S_n$. The field F is called the field of *symmetric functions* in the x_i. The reason for this name is that if $f(x_1, \ldots, x_n)/g(x_1, \ldots, x_n) \in F$, then

$$f(x_{\sigma(1)}, \ldots, x_{\sigma(n)})/g(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = f(x_1, \ldots, x_n)/g(x_1, \ldots, x_n)$$

for all $\sigma \in S_n$. Let

$$\begin{aligned} s_1 &= x_1 + x_2 + \cdots + x_n, \\ s_2 &= x_1x_2 + x_1x_3 + \cdots + x_{n-1}x_n, \\ &\vdots \\ s_n &= x_1x_2 \cdots x_n. \end{aligned}$$

The polynomial s_i is called the ith *elementary symmetric function.* We see that each $s_i \in F$, so $k(s_1, \ldots, s_n) \subseteq F$. Note that

$$(t - x_1) \cdots (t - x_n) = t^n - s_1t^{n-1} + s_2t^{n-2} + \cdots + (-1)^n s_n.$$

From this fact, we shall see in Section 3 that $F = k(s_1, \ldots, s_n)$. This means that every symmetric function in the x_i is a rational function in the elementary symmetric functions.

Problems

1. Show that the only automorphism of $\mathbb{Q}$ is the identity.

2. Show that the only automorphism of $\mathbb{R}$ is the identity. (Hint: If σ is an automorphism, show that $\sigma|_{\mathbb{Q}} = \text{id}$, and if $a > 0$, then $\sigma(a) > 0$. It is an interesting fact that there are infinitely many automorphisms of $\mathbb{C}$, even though $[\mathbb{C} : \mathbb{R}] = 2$. Why is this fact not a contradiction to this problem?)

3. Show that the six functions given in Example 2.21 extend to $\mathbb{Q}$-automorphisms of $\mathbb{Q}(\sqrt[3]{2}, \omega)$.

4. Let B be an integral domain with quotient field F. If $\sigma : B \to B$ is a ring automorphism, show that σ induces a ring automorphism $\sigma' : F \to F$ defined by $\sigma'(a/b) = \sigma(a)/\sigma(b)$ if $a, b \in B$ with $b \neq 0$.

5. Let $K = k(x_1, \ldots, x_n)$ be the field of rational functions in n variables over a field k. Show that the definition

$$\sigma\left(\frac{f(x_1, \ldots, x_n)}{g(x_1, \ldots, x_n)}\right) = \frac{f(x_{\sigma(1)}, \ldots, x_{\sigma(n)})}{g(x_{\sigma(1)}, \ldots, x_{\sigma(n)})}$$

makes a permutation $\sigma \in S_n$ into a field automorphism of K. (Hint: The previous problem along with Problem 1.6 may help some.)

6. Let F be a field of characteristic not 2, and let K be an extension of F with $[K : F] = 2$. Show that $K = F(\sqrt{a})$ for some $a \in F$; that is, show that $K = F(\alpha)$ with $\alpha^2 = a \in F$. Moreover, show that K is Galois over F.

7. Let $F = \mathbb{F}_2$ and $K = F(\alpha)$, where α is a root of $1 + x + x^2$. Show that the function $\sigma : K \to K$ given by $\sigma(a + b\alpha) = a + b + b\alpha$ for $a, b \in F$ is an F-automorphism of K.

8. Suppose that $a \in \mathbb{C}$ is algebraic over $\mathbb{Q}$ with $p(x) = \min(\mathbb{Q}, a)$, and let b be any root in $\mathbb{C}$ of p. Show that the map $\sigma : \mathbb{Q}(a) \to \mathbb{C}$ given by $\sigma(f(a)) = f(b)$ is a well-defined $\mathbb{Q}$-homomorphism.

9. Show that the complex numbers $i\sqrt{3}$ and $1 + i\sqrt{3}$ are roots of $f(x) = x^4 - 2x^3 + 7x^2 - 6x + 12$. Let K be the field generated by $\mathbb{Q}$ and the roots of f. Is there an automorphism σ of K with $\sigma(i\sqrt{3}) = 1 + i\sqrt{3}$?

10. Determine whether the following fields are Galois over $\mathbb{Q}$.

    (a) $\mathbb{Q}(\omega)$, where $\omega = \exp(2\pi i/3)$.

    (b) $\mathbb{Q}(\sqrt[4]{2})$.

    (c) $\mathbb{Q}(\sqrt{5}, \sqrt{7})$. (Hint: The previous section has a problem that might be relevant.)

11. Prove or disprove the following assertion and its converse: If $F \subseteq L \subseteq K$ are fields with K/L and L/F Galois, then K/F is Galois.

12. *Galois connections.* The relationship given in Corollary 2.10 between the set of intermediate fields of a Galois extension and the set of subgroups of its Galois group appears in other situations, so we study it here. We first need a definition. If S is a set, a relation $\leq$ on S is called a *partial order* on S provided that $a \leq a$ for all $a \in S$; if $a \leq b$ and $b \leq a$, then $a = b$; and if $a \leq b$ and $b \leq c$, then $a \leq c$. Let S and T be sets with partial orders $\leq_S$ and $\leq_T$, respectively. Suppose that there are functions $f : S \to T$ and $g : T \to S$ such that (i) if $s_1 \leq_S s_2$, then $f(s_2) \leq_T f(s_1)$, (ii) if $t_1 \leq_T t_2$, then $f(t_2) \leq_S f(t_1)$, and (iii) $s \leq_S g(f(s))$ and $t \leq_T f(g(t))$ for all $s \in S$ and $t \in T$. Prove that there is a 1–1 order reversing correspondence between the image of g and the image of f, given by $s \mapsto f(s)$, whose inverse is $t \mapsto g(t)$.

13. Let k be a field, and let $K = k(x)$ be the rational function field in one variable over k. Let σ and τ be the automorphisms of K defined by $\sigma(f(x)/g(x)) = f(1/x)/g(1/x)$ and $\tau(f(x)/g(x)) = f(1-x)/g(1-x)$, respectively. Determine the fixed field F of $\{\sigma, \tau\}$, and determine $\operatorname{Gal}(K/F)$. Find an $h \in F$ so that $F = k(h)$.

14. Let k be a field, and let $K = k(x)$ be the rational function field in one variable over k. If $u \in K$, show that $K = k(u)$ if and only if $u = (ax + b)/(cx + d)$ for some $a, b, c, d \in k$ with $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \neq 0$. (Hint: See the example before Proposition 1.15.)

15. Use the previous problem to show that any invertible 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ determines an element of $\operatorname{Gal}(k(x)/k)$ with $x \mapsto (ax + b)/(cx + d)$. Moreover, show that every element of $\operatorname{Gal}(k(x)/k)$ is given by such a formula. Show that the map from the set of invertible 2×2 matrices over k to $\operatorname{Gal}(k(x)/k)$ given by $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \varphi$, where $\varphi(x) = (ax + b)/(cx + d)$, is a group homomorphism. Determine the kernel to show that $\operatorname{Gal}(k(x)/k) \cong \operatorname{PGL}_2(k)$, the group of invertible 2×2 matrices over k modulo the scalar matrices.
(This group is the *projective general linear group over* k of 2×2 matrices.)

16. Let $k = \mathbb{R}$, and let A be the matrix $\begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}$ given by rotating the plane around the origin by 120°. Using the previous problem, show that A determines a subgroup of $\operatorname{Gal}(k(x)/k)$ of order 3. Let F be the fixed field. Show that $k(x)/F$ is Galois, find a u so that $F = k(u)$, find the minimal polynomial $\min(F, x)$, and find all the roots of this polynomial.

17. Let $k = \mathbb{F}_p$, and let $k(x)$ be the rational function field in one variable over k. Define $\varphi : k(x) \to k(x)$ by $\varphi(x) = x + 1$. Show that φ has finite order in $\text{Gal}(k(x)/k)$. Determine this order, find a u so that $k(u)$ is the fixed field of φ, determine the minimal polynomial over $k(u)$ of x, and find all the roots of this minimal polynomial.

18. Let k be a field of characteristic $p > 0$, and let $a \in k$. Let $f(x) = x^p - a^{p-1}x$. Show that f is fixed by the automorphism φ of $k(x)$ defined by $\varphi(f(x)/g(x)) = f(x+a)/g(x+a)$ for any $f(x), g(x) \in k[x]$. Show that $k(f)$ is the fixed field of φ.

19. Prove that $(t - x_1) \cdots (t - x_n) = t^n - s_1 t^{n-1} + \cdots + (-1)^n s_n$, as we claimed in Example 2.22.

3 Normal Extensions

In the last section, we saw that there are two ways for the field extension $F(a)/F$ to fail to be Galois: if $\min(F, a)$ does not have all its roots in $F(a)$ or if $\min(F, a)$ has repeated roots. The next two sections investigate these two situations. In this section, we investigate the case when $F(a)$ contains all the roots of $p(x)$ and what this question means for general algebraic extensions. We begin with a result that in the case of polynomials over $\mathbb{R}$ should be familiar.

Lemma 3.1 *Let $f(x) \in F[x]$ and $\alpha \in F$. Then α is a root of f if and only if $x - \alpha$ divides f. Furthermore, f has at most* $\deg(f)$ *roots in any extension field of F.*

Proof. By the division algorithm, $f(x) = q(x) \cdot (x - \alpha) + r(x)$ for some $q(x)$ and $r(x)$ with $r(x) = 0$ or $\deg(r) < \deg(x - \alpha)$. In either case, we see that $r(x) = r$ is a constant. But $f(\alpha) = r$, so $f(\alpha) = 0$ if and only if $x - \alpha$ divides $f(x)$.

For the second part, we argue by induction on $n = \deg(f)$. If $n = 1$, then $f(x) = ax + b$ for some $a, b \in F$. The only root of f is $-b/a$, so the result is true if $n = 1$. Assume that any polynomial over an extension field of F of degree $n - 1$ has at most $n - 1$ roots in any extension field K of F. If $f(x)$ has no roots in K, then we are done. If instead $\alpha \in K$ is a root of f, then $f(x) = (x - \alpha) \cdot g(x)$ for some $g(x) \in K[x]$ by the first part of the lemma. Since $g(x)$ has degree $n - 1$, by induction g has at most $n - 1$ roots in K. The roots of f consist of α together with the roots of g. Thus, f has at most n roots. $\square$

Definition 3.2 *If K is an extension field of F and if $f(x) \in F[x]$, then f splits over K if $f(x) = a \prod_i (x - \alpha_i) \in K[x]$ for some $\alpha_1, \ldots, \alpha_n \in K$ and*

$a \in F$. In other words, f splits over K if f factors completely into linear factors in $K[x]$.

In order to talk about roots of a given polynomial, we need to have extension fields that contain the roots of the polynomial. The next theorem shows that for any $f(x) \in F[x]$, there is a finite extension of F over which f splits. We use a generalization of the construction of Example 1.6 to construct a field containing roots of a given polynomial.

Theorem 3.3 *Let $f(x) \in F[x]$ have degree n. There is an extension field K of F with $[K : F] \leq n$ such that K contains a root of f. In addition, there is a field L containing F with $[L : F] \leq n!$ such that f splits over L.*

Proof. Let $p(x)$ be an irreducible factor of $f(x)$ in $F[x]$, and let K be the field $F[x]/(p(x))$. Then F is isomorphic to a subfield of K; namely the map $\varphi : F \longrightarrow K$ given by $\varphi(a) = a + (p(x))$ is an injection of fields. We will view $F \subseteq K$ by replacing F with $\varphi(F)$. If $\alpha = x + (p(x)) \in K$, then $p(\alpha) = p(x) + (p(x)) = 0 + (p(x))$. Thus, α is a root of p in K; therefore, α is a root of f. Since $[K : F] = \deg(p) \leq n$, this proves the first part of the theorem.

For the second part, we use induction on n. By the first part, there is a field $K \supseteq F$ with $[K : F] \leq n$ such that K contains a root α of $f(x)$, say $f(x) = (x - \alpha) \cdot g(x)$ with $g(x) \in K[x]$. By induction, there is a field $L \supseteq K$ with $[L : K] \leq (n-1)!$ such that g splits over L. But then f splits over L and $[L : F] = [L : K] \cdot [K : F] \leq (n-1)! \cdot n = n!$. □

Definition 3.4 *Let K be an extension field of F and let $f(x) \in F[x]$.*

1. *If $f(x) \in F[x]$, then K is a splitting field of f over F if f splits over K and $K = F(\alpha_1, \ldots, \alpha_n)$, where $\alpha_1, \ldots, \alpha_n$ are the roots of f.*

2. *If S is a set of nonconstant polynomials over F, then K is a splitting field of S over F if each $f \in S$ splits over K and $K = F(X)$, where X is the set of all roots of all $f \in S$.*

Intuitively, a splitting field for a set S of polynomials is a minimal field extension over which each $f \in S$ splits. This is made more concrete in Problem 2.

Theorem 3.3 yields immediately the existence of splitting fields for a finite set of polynomials.

Corollary 3.5 *If $f_1(x), \ldots, f_n(x) \in F[x]$, then there is a splitting field for $\{f_1, \ldots, f_n\}$ over F.*

Proof. Suppose that $f_1, \ldots, f_n \in F[x]$. Note that a splitting field of $\{f_1, \ldots, f_n\}$ is the same as a splitting field of the product $f_1 \cdots f_n$. If $f = f_1 \cdots f_n$, then by Theorem 3.3, there is a field $L \supseteq F$ such that f

splits over L. Let $\alpha_1, \ldots, \alpha_n \in L$ be the roots of f. Then $F(\alpha_1, \ldots, \alpha_n)$ is a splitting field for f over F. □

Example 3.6 The field $\mathbb{Q}(\omega, \sqrt[3]{2})$ is a splitting field for $x^3 - 2$ over $\mathbb{Q}$, since we have seen in Example 2.21 that this field is also the field generated by the three roots of $x^3 - 2$ over $\mathbb{Q}$. The complex field $\mathbb{C}$ is a splitting field over $\mathbb{R}$ for $x^2 + 1$, since $\mathbb{C} = \mathbb{R}(i, -i)$ is generated by $\mathbb{R}$ and the roots of $x^2 + 1$. In general, if F is a field and $a \in F$, then the field $F(\sqrt{a})$ is a splitting field for $x^2 - a$ over F.

Example 3.7 Let $F = \mathbb{F}_2$ and $K = F[x]/(1 + x + x^2) \cong F(\alpha)$, where α is a root of $1 + x + x^2$. Then $1 + x + x^2$ factors as $(x - \alpha)(x - (\alpha + 1))$ over K, so K is a splitting field of $1 + x + x^2$.

We will show that splitting fields are unique up to isomorphism. From this fact, the next corollary would follow from Theorem 3.3. However, we give a different proof so that we can use it in the next example.

Corollary 3.8 *Let F be a field and let $f(x) \in F[x]$ be a polynomial of degree n. If K is a splitting field of f over F, then $[K : F] \le n!$.*

Proof. We prove this by induction on $n = \deg(f)$. If $n = 1$, then the result is clear. Suppose that $n > 1$ and that the result is true for polynomials of degree $n - 1$. Let K be a splitting field of f over F, and let a be a root of f in K. Then $[F(a) : F] \le n$, since $\min(F, a)$ divides f. If $f(x) = (x - a)g(x)$, then $\deg(g) = n - 1$ and K is the splitting field of g over $F(a)$. By induction, $[K : F(a)] \le (n - 1)!$ by Theorem 3.3, so

$$\begin{aligned} [K : F] &= [F(a) : F] \cdot [K : F(a)] \\ &\le n \cdot (n - 1)! = n!. \end{aligned}$$

This proves the corollary. □

Example 3.9 Let k be a field, and let $K = k(x_1, x_2, \ldots, x_n)$ be the rational function field in n variables over k. We view the symmetric group S_n as a subgroup of $\mathrm{Aut}(K)$ by defining

$$\sigma\left(\frac{f(x_1, \ldots, x_n)}{g(x_1, \ldots, x_n)}\right) = \frac{f(x_{\sigma(1)}, \ldots, x_{\sigma(n)})}{g(x_{\sigma(1)}, \ldots, x_{\sigma(n)})}$$

for $\sigma \in S_n$, as in Example 2.22. Let $F = \mathcal{F}(S_n)$, the field of symmetric functions in the x_i. Then $S_n = \mathrm{Gal}(K/F)$ by Proposition 2.14, so $[K : F] = |S_n| = n!$. We wish to determine F. Let $s_1, s_2, \ldots, s_n$ be the elementary symmetric functions in the x_i; that is,

$$s_1 = x_1 + x_2 + \cdots + x_n,$$

$$s_2 = \sum_{i \neq j} x_i x_j,$$

$$\vdots$$

$$s_n = x_1 x_2 \cdots x_n.$$

Then $k(s_1, s_2, \ldots, s_n) \subseteq F$. We claim that $F = k(s_1, \ldots, s_n)$. To show this, we use the concept of splitting fields. Let

$$f(t) = t^n - s_1 t^{n-1} + \cdots + (-1)^n s_n \in k(s_1, s_2, \ldots, s_n)[t].$$

Then $f(t) = (t-x_1)\cdots(t-x_n)$ in $K[x]$, which can be seen by expanding this product. Since K is generated over k by the x_i, we see that K is a splitting field for $f(t)$ over $k(s_1, s_2, \ldots, s_n)$. We know that $[K : F] = |S_n| = n!$, and so $[K : k(s_1, s_2, \ldots, s_n)] \geq n!$. However, $[K : k(s_1, s_2, \ldots, s_n)] \leq n!$ by Corollary 3.8. Therefore, $[K : k(s_1, s_2, \ldots, s_n)] = [K : F]$. This forces $F = k(s_1, s_2, \ldots, s_n)$. Therefore, any symmetric function can be written in terms of the elementary symmetric functions. In fact, every symmetric polynomial can be written as a polynomial in the elementary symmetric functions (see Problem 17).

Algebraic closures

We have proved the existence of splitting fields for finite sets of polynomials. What about infinite sets? Suppose that K is a splitting field over F of the set of all nonconstant polynomials over F. We do not know yet that such a field exists, but we will show it does exist. Let L be an algebraic extension of K. If $a \in L$, then a is algebraic over F by Theorem 1.24, since K is algebraic over F. Let $f(x) = \min(F, a)$. Then f splits over K; hence, $a \in K$. Thus, $L = K$. This proves that K has no algebraic extensions. The existence of such a field will imply the existence of splitting fields of an arbitrary set of polynomials. Moreover, given K, we shall see that any algebraic extension of F is isomorphic to a subfield of K. This will allow us to view all algebraic extensions of F as subfields of K.

We first give some equivalent conditions for such a field.

Lemma 3.10 *If K is a field, then the following statements are equivalent:*

1. *There are no algebraic extensions of K other than K itself.*
2. *There are no finite extensions of K other than K itself.*
3. *If L is a field extension of K, then $K = \{a \in L : a$ is algebraic over $K\}$.*
4. *Every $f(x) \in K[x]$ splits over K.*
5. *Every $f(x) \in K[x]$ has a root in K.*

6. *Every irreducible polynomial over K has degree* 1.

Proof. (1) $\Rightarrow$ (2): This is clear, since any finite extension of F is an algebraic extension of F.

(2) $\Rightarrow$ (3): Let $a \in L$ be algebraic over K. Then $K(a)$ is a finite extension of K, so $K(a) = K$. Thus, $a \in K$.

(3) $\Rightarrow$ (4): Let $f(x) \in K[x]$, and let L be a splitting field of f over K. Since L is algebraic over K, statement 3 shows that $L = K$; that is, f splits over K.

(4) $\Rightarrow$ (5): This is clear.

(5) $\Rightarrow$ (6): Let $f(x) \in K[x]$ be irreducible. By statement 5, f has a root in K, so f has a linear factor. Since f is irreducible, this means f itself is linear, so $\deg(f) = 1$.

(6) $\Rightarrow$ (1): Let L be an algebraic extension of K. Take $a \in L$ and let $p(x) = \min(K, a)$. By statement 6, the degree of p is 1, so $[K(a) : K] = 1$. Thus, $a \in K$, so $L = K$. □

Definition 3.11 *If K satisfies the equivalent conditions of Lemma* 3.10, *then K is said to be algebraically closed. If K is an algebraic extension of F and is algebraically closed, then K is said to be an algebraic closure of F.*

Example 3.12 The complex field $\mathbb{C}$ is algebraically closed. This fact is usually referred to as the *fundamental theorem of algebra*, and it will be proved in Section 5. If

$$\mathbb{A} = \{a \in \mathbb{C} : a \text{ is algebraic over } \mathbb{Q}\},$$

then it is not hard to prove that $\mathbb{A}$ is algebraically closed by using that $\mathbb{C}$ is algebraically closed; see Problem 4b. Furthermore, $\mathbb{C}$ is an algebraic closure of $\mathbb{R}$, and $\mathbb{A}$ is an algebraic closure of $\mathbb{Q}$. However, $\mathbb{C}$ is not an algebraic closure of $\mathbb{Q}$ since $\mathbb{C}$ is not algebraic over $\mathbb{Q}$.

We wish to prove the existence of an algebraic closure of an arbitrary field F and to prove the existence of a splitting field for an arbitrary set of polynomials. In order to do this, we will use a Zorn's lemma argument. The next lemma is needed for technical reasons in the proof of the existence of an algebraic closure.

Lemma 3.13 *If K/F is algebraic, then $|K| \le \max\{|F|, |\mathbb{N}|\}$.*

Proof. In this proof, we require some facts of cardinal arithmetic, facts that can be found in Proposition 2.1 in Appendix B. If $a \in K$, pick a labeling $a_1, \dots, a_n$ of the roots of $\min(F, a)$ in K. If $\mathcal{M}$ is the set of all monic polynomials over F, define $f : K \longrightarrow \mathcal{M} \times \mathbb{N}$ by $f(a) = (p(x), r)$ if $p(x) = \min(F, a)$ and $a = a_r$. This map is clearly injective, so

$$|K| \le |\mathcal{M} \times \mathbb{N}| = \max\{|\mathcal{M}|, |\mathbb{N}|\}.$$

We will be done by showing that $|\mathcal{M}| \leq \max\{|F|, |\mathbb{N}|\}$. For this, if $\mathcal{M}_n$ is the set of monic polynomials over F of degree n, then $|\mathcal{M}_n| = |F^n|$, since the map $(a_0, \dots, a_{n-1}) \mapsto x^n + \sum_{i=0}^{n-1} a_i x^i$ is a bijection between F^n and $\mathcal{M}_n$. If F is finite, then $|F^n| = |F|^n$ is finite, and if F is infinite, then $|F^n| = |F|$. Therefore, since $\mathcal{M}$ is the union of the disjoint sets $\mathcal{M}_n$, we have $|\mathcal{M}| = |\bigcup_n \mathcal{M}_n| = \max\{|F|, |\mathbb{N}|\}$. □

Theorem 3.14 *Let F be a field. Then F has an algebraic closure.*

Proof. Let S be a set containing F with $|S| > \max\{|F|, |\mathbb{N}|\}$. Let $\mathcal{A}$ be the set of all algebraic extension fields of F inside S. Then $\mathcal{A}$ is ordered by defining $K \leq L$ if L is an extension field of K. By Zorn's lemma, there is a maximal element M of $\mathcal{A}$. We claim that M is an algebraic closure of M. To show that M is algebraically closed, let L be an algebraic extension of M. By Lemma 3.13,

$$|L| \leq \max\{|M|, |\mathbb{N}|\} \leq \{|F|, |\mathbb{N}|\} < |S|.$$

Thus, there is a function $f : L \longrightarrow S$ with $f|_M = \text{id}$. By defining $+$ and $\cdot$ on $f(L)$ by $f(a) + f(b) = f(a+b)$ and $f(a) \cdot f(b) = f(ab)$, we see that $f(L)$ is a field extension of M and f is a field homomorphism. Maximality of M shows that $f(L) = M$, so $L = M$. Thus, M is algebraically closed. Since M is algebraic over F, we see that M is an algebraic closure of F. □

The existence of an algebraic closure yields immediately the existence of a splitting field for an arbitrary set of nonconstant polynomials.

Corollary 3.15 *Let S be a set of nonconstant polynomials over F. Then S has a splitting field over F.*

Proof. Let K be an algebraic closure of F. Then each $f(x) \in S$ splits over K. Let X be the set of roots of all $f \in S$. Then $F(X) \subseteq K$ is a splitting field for S over F, since each f splits over $F(X)$ and this field is generated by the roots of all the polynomials from S. □

To emphasize a useful interpretation of an algebraic closure, we record the following easy consequence of the existence of arbitrary splitting fields.

Corollary 3.16 *If F is a field, then the splitting field of the set of all nonconstant polynomials over F is an algebraic closure of F.*

Now that we have the existence of a splitting field for any set of nonconstant polynomials, what can we say about such fields? Can we have many different splitting fields, up to isomorphism? The answer is no; the next lemma is the first step in showing this.

The following fact is used in the lemma below and in a number of other places. If $\sigma : F \to F'$ is a field homomorphism, then there is an induced

ring homomorphism $F[x] \longrightarrow F'[x]$, which we also denote by σ, given by $\sigma\left(\sum a_i x^i\right) = \sum \sigma(a_i)x^i$. It is an easy calculation to show that σ does indeed induce a ring homomorphism on $F[x]$. If $f(x) = (x - a_1)\cdots(x - a_n) \in F[x]$, then the preservation of polynomial multiplication shows that $\sigma(f(x)) = (x - \sigma(a_1))\cdots(x - \sigma(a_n))$. This relationship between σ and factorization of polynomials will help us to study splitting fields.

Lemma 3.17 *Let $\sigma : F \longrightarrow F'$ be a field isomorphism. Let $f(x) \in F[x]$ be irreducible, let α be a root of f in some extension field K of F, and let α' be a root of $\sigma(f)$ in some extension K' of F'. Then there is an isomorphism $\tau : F(\alpha) \longrightarrow F'(\alpha')$ with $\tau(\alpha) = \alpha'$ and $\tau|_F = \sigma$.*

Proof. Since f is irreducible and $f(\alpha) = 0$, the minimal polynomial of α over F is a constant multiple of f. Thus, f and $\min(F, \alpha)$ generate the same principal ideal in $F[x]$. We then have an F-isomorphism $\varphi : F[x]/(f(x)) \longrightarrow F(\alpha)$ given by $\varphi(g(x) + (f(x))) = g(\alpha)$ and an F'-isomorphism $\psi : F'[x]/(f'(x)) \longrightarrow F'(\alpha')$ given by $\psi(g(x) + (f'(x))) = g(\alpha')$. Since $\sigma(f) = f'$, the map $\nu(g(x) + (f(x))) = \sigma(g(x)) + (f'(x))$ gives a well-defined isomorphism $\nu : F[x]/(f(x)) \longrightarrow F'[x]/(f'(x))$ which extends σ. We have the following sequence of field isomorphisms:

$$F(\alpha) \xrightarrow{\varphi^{-1}} F[x]/(f(x)) \xrightarrow{\nu} F'[x]/(f'(x)) \xrightarrow{\psi} F'(\alpha').$$

Therefore, the composition $\varphi^{-1} \circ \nu \circ \psi : F(\alpha) \longrightarrow F(\alpha')$ is an isomorphism extending σ on F with $\alpha \mapsto x + (f(x)) \mapsto x + (f'(x)) \mapsto \alpha'$. □

Lemma 3.18 *Let $\sigma : F \longrightarrow F'$ be a field isomorphism, let K be a field extension of F, and let K' be a field extension of F'. Suppose that K is a splitting field of $\{f_i\}$ over F and that $\tau : K \longrightarrow K'$ is a homomorphism with $\tau|_F = \sigma$. If $f_i' = \sigma(f_i)$, then $\tau(K)$ is a splitting field of $\{f_i'\}$ over F'.*

Proof. Because K is a splitting field of a set $\{f_i\}$ of polynomials over F, given f_i there are $a, \alpha_1, \ldots, \alpha_n \in K$ with $f_i(x) = a\prod_j(x - \alpha_j)$. Therefore, $\tau(f_i(x)) = \tau(a)\prod_j(x - \tau(\alpha_j))$. Hence, each $f_i' = \sigma(f_i) = \tau(f_i)$ splits over $\tau(K)$. Since K is generated over F by the roots of the f_i, the field $\tau(K)$ is generated over F' by the images of the roots of the f_i; that is, $\tau(K)$ is generated over F' by the roots of the f_i'. Thus, $\tau(K)$ is a splitting field over F' for $\{f_i'\}$. □

The next theorem, the isomorphism extension theorem, is one of the most important results of Galois theory. It proves the uniqueness of splitting fields, although its main use is in constructing automorphisms of a field, and thus for calculating the Galois group of a field extension. Before proving it, we give a proof of the case of splitting fields of a single polynomial. While the full version certainly includes this case, we give a proof of this special case for a few reasons: The proof of this special case is easy and the

argument is useful to see, many applications of the theorem require only this case, and the full proof uses a Zorn's lemma argument and is not very intuitive.

Theorem 3.19 *Let $\sigma : F \to F'$ be a field isomorphism, let $f(x) \in F[x]$, and let $\sigma(f) \in F'[x]$ be the corresponding polynomial over F'. Let K be the splitting field of f over F, and let K' be the splitting field of $\sigma(f)$ over F'. Then there is an isomorphism $\tau : K \longrightarrow K'$ with $\tau|_F = \sigma$. Furthermore, if $\alpha \in K$ and if α' is any root of $\sigma(\min(F, \alpha))$ in K', then τ can be chosen so that $\tau(\alpha) = \alpha'$.*

Proof. We prove this by induction on $n = [K : F]$. If $n = 1$, then f splits over F, and the result is trivial in this case. So, suppose that $n > 1$ and that the result is true for splitting fields of degree less than n. If f splits over F, then the result is clear. If not, let $p(x)$ be a nonlinear irreducible factor of $f(x)$, let α be a root of p, and let α' be a root of $\sigma(p)$. Set $L = F(\alpha)$ and $L' = F(\alpha')$. Then $[L : F] > 1$, so $[K : L] < n$. By Lemma 3.17, there is a field isomorphism $\rho : L \to L'$ with $\rho(\alpha) = \alpha'$. Since K is the splitting field over L for $f(x)$ and K' is the splitting field over L' for $\sigma(f)$, by induction the isomorphism τ extends to an isomorphism $\tau : K \to K'$. The isomorphism τ is then an extension of σ (and ρ), and $\tau(\alpha) = \rho(\alpha) = \alpha'$. □

Theorem 3.20 (Isomorphism Extension Theorem) *Let $\sigma : F \longrightarrow F'$ be a field isomorphism. Let $S = \{f_i(x)\}$ be a set of polynomials over F, and let $S' = \{\sigma(f_i)\}$ be the corresponding set over F'. Let K be a splitting field for S over F, and let K' be a splitting field for S' over F'. Then there is an isomorphism $\tau : K \longrightarrow K'$ with $\tau|_F = \sigma$. Furthermore, if $\alpha \in K$ and α' is any root of $\sigma(\min(F, \alpha))$ in K', then τ can be chosen so that $\tau(\alpha) = \alpha'$.*

Proof. We prove this with a Zorn's lemma argument. Let $\mathcal{S}$ be the set of all pairs (L, φ) such that L is a subfield of K and $\varphi : L \longrightarrow K'$ is a homomorphism extending σ. This set is nonempty since $(F, \sigma) \in \mathcal{S}$. Furthermore, $\mathcal{S}$ is partially ordered by defining $(L, \varphi) \leq (L', \varphi')$ if $L \subseteq L'$ and $\varphi'|_L = \varphi$. Let $\{(L_i, \varphi_i)\}$ be a chain in $\mathcal{S}$. If $L = \bigcup_i L_i$ and $\varphi : L \longrightarrow K'$ is defined by $\varphi(a) = \varphi_i(a)$ if $a \in L_i$, then it is not hard to see that L is a field extension of all the L_i and φ is a homomorphism extending σ. Thus, (L, φ) is an upper bound in $\mathcal{S}$ for this chain. Therefore, by Zorn's lemma there is a maximal element (M, τ) in $\mathcal{S}$. We claim that $M = K$ and $\tau(M) = K'$. If $M \neq K$, then there is an $f \in S$ that does not split over M. Let $\alpha \in K$ be a root of f that is not in M, and let $p(x) = \min(F, a)$. Set $p' = \sigma(p) \in F'[x]$ and let $\alpha' \in K'$ be a root of p'. Such an α' exists since p' divides f' and f' splits over K'. By Lemma 3.17, there is a $\rho : M(\alpha) \longrightarrow \tau(M)(\alpha')$ that extends τ. Then $(M(\alpha), \rho) \in S$ is larger than (M, τ), a contradiction to the maximality of (M, τ). This proves that $M = K$. The equality $\tau(K) = K'$

follows immediately from Lemma 3.18, since $\tau(K) \subseteq K'$ is a splitting field for S' over F'. □

Corollary 3.21 *Let F be a field, and let S be a subset of $F[x]$. Any two splitting fields of S over F are F-isomorphic. In particular, any two algebraic closures of F are F-isomorphic.*

Proof. For the proof of the first statement, the isomorphism extension theorem gives an isomorphism extending id on F between any two splitting fields of S. The second statement follows from the first, since any algebraic closure of F is a splitting field of the set of all nonconstant polynomials in $F[x]$. □

As a corollary to the existence and uniqueness of algebraic closures, we can prove that any algebraic extension of a field F can be viewed as living inside a fixed algebraic closure of F.

Corollary 3.22 *Let F be a field, and let N be an algebraic closure of F. If K is an algebraic extension of F, then K is isomorphic to a subfield of N.*

Proof. Let M be an algebraic closure of K. By Theorem 1.24, M is algebraic over F; hence, M is also an algebraic closure of F. Therefore, by the previous corollary, $M \cong N$. If $f : M \to N$ is an F-isomorphism, then $f(K)$ is a subfield of N isomorphic to K. □

We now go into more detail about splitting fields. One question we will address is the following. If K is the splitting field of a set S of polynomials over F, can we determine all of the polynomials in $F[x]$ that split over K? Also, can we give a more intrinsic characterization of K, one that does not refer to the set S? The answer to both questions is yes and is found in Proposition 3.28.

Definition 3.23 *If K is a field extension of F, then K is normal over F if K is a splitting field of a set of polynomials over F.*

Example 3.24 If $[K : F] = 2$, then K is normal over F. For, if $a \in K - F$, then $K = F(a)$, since $[K : F] = 2$. If $p(x) = \min(F, a)$, then p has one root in K; hence, since $\deg(p) = 2$, this polynomial factors over K. Because K is generated over F by the roots of $p(x)$, we see that K is a splitting field for $p(x)$ over F.

Example 3.25 If $F \subseteq L \subseteq K$ are fields such that K/F is normal, then K/L is normal. This is true because if K is the splitting field over F of a set of polynomials $S \subseteq F[x]$, then K is generated over F by the roots of the polynomials in S. Consequently, K is generated by the roots as an

extension of L, so K is a splitting field of S over L, and so K is normal over L.

Example 3.26 The field $\mathbb{Q}(\omega, \sqrt[3]{2})$ is normal over $\mathbb{Q}$, since it is the splitting field of $x^3 - 2$ over $\mathbb{Q}$. Similarly, if $i = \sqrt{-1}$, then $\mathbb{Q}(\sqrt[4]{2}, i)$ is normal over $\mathbb{Q}$, since it is the splitting field of $x^4 - 2$ over $\mathbb{Q}$. The subfield $\mathbb{Q}(i)$ is also normal over $\mathbb{Q}$, as it is the splitting field of $x^2 + 1$ over $\mathbb{Q}$. However, the subfield $\mathbb{Q}(\sqrt[4]{2})$ is not normal over $\mathbb{Q}$. At this point, we do not have an effective way of showing $\mathbb{Q}(\sqrt[4]{2})/\mathbb{Q}$ is not normal, for we would have to show that there is no polynomial $f \in \mathbb{Q}(x)$ whose roots generate $\mathbb{Q}(\sqrt[4]{2})$. It is clear that $\min(\mathbb{Q}, \sqrt[4]{2})$ does not split over $\mathbb{Q}(\sqrt[4]{2})$, which will be enough to show that $\mathbb{Q}(\sqrt[4]{2})$ is not normal over $\mathbb{Q}$ once we prove Proposition 3.28.

Example 3.27 Let F be a field of characteristic $p > 0$, and suppose that $K = F(a_1, \ldots, a_n)$ with $a_i^p \in F$ for each i. Then we show that K is normal over F. The minimal polynomial of a_i divides $x^p - a_i^p$, which factors completely over K as $x^p - a_i^p = (x - a_i)^p$; hence, $\min(F, a_i)$ splits over K. Thus, K is the splitting field of $\{\min(F, a_i) : 1 \le i \le n\}$ over F. Note that each $\min(F, a_i)$ has only one distinct root, and any F-automorphism of K is determined by its action on the generators $a_1, \ldots, a_n$, so $\mathrm{Gal}(K/F) = \{\mathrm{id}\}$. For instance, if $k(x_1, \ldots, x_n)$ is the rational function field in n variables over a field k of characteristic p, then $k(x_1, \ldots, x_n)/k(x_1^p, \ldots, x_n^p)$ is a normal extension.

If K is the splitting field over F of a set of polynomials $S \subseteq F[x]$, then each polynomial in S splits over K. However, K can be viewed as a splitting field in other ways, as the following proposition shows.

Proposition 3.28 *If K is algebraic over F, then the following statements are equivalent:*

1. *The field K is normal over F.*
2. *If M is an algebraic closure of K and if $\tau : K \longrightarrow M$ is an F-homomorphism, then $\tau(K) = K$.*
3. *If $F \subseteq L \subseteq K \subseteq N$ are fields and if $\sigma : L \longrightarrow N$ is an F-homomorphism, then $\sigma(L) \subseteq K$, and there is a $\tau \in \mathrm{Gal}(K/F)$ with $\tau|_L = \sigma$.*
4. *For any irreducible $f(x) \in F[x]$, if f has a root in K, then f splits over K.*

Proof. (1) $\Rightarrow$ (2): Let M be an algebraic closure of K, and let $\tau : K \longrightarrow M$ be an F-homomorphism. If K is the splitting field for $S \subseteq F[x]$ over F, then so is $\tau(K) \subseteq M$ by Lemma 3.17. Since K and $\tau(K)$ are generated over F by the same set of roots, $K = \tau(K)$.

(2) $\Rightarrow$ (3): Suppose that $F \subseteq L \subseteq K \subseteq N$ are fields and that $\sigma : L \longrightarrow N$ is an F-homomorphism. Since $L \subseteq K$, the extension L/F is algebraic, and so $\sigma(L) \subseteq N$ is algebraic over F. Let M' be the algebraic closure of F in N and let M be an algebraic closure of M'. Then M is also an algebraic closure of K. By the isomorphism extension theorem, there is an extension $\rho : M \longrightarrow M$ with $\rho|_L = \sigma$. Let $\tau = \rho|_K$. By condition 2 we have $\tau(K) = K$, so $\sigma(L) = \tau(L) \subseteq \tau(K) = K$. Thus, $\tau \in \text{Gal}(K/F)$.

(3) $\Rightarrow$ (4): Let $f(x) \in F[x]$ be irreducible over F, and let $\alpha \in K$ be a root of f. Let $L = F(\alpha) \subseteq K$ and let N be an algebraic closure of K. If $\beta \in M$ is any root of f, then there is an F-homomorphism $\sigma : L \longrightarrow M$ given by $g(\alpha) \mapsto g(\beta)$. By condition 3, $\sigma(L) \subseteq K$, so $\beta \in K$. Hence, all roots of f lie in K, so f splits over K.

(4) $\Rightarrow$ (1): Condition 4 shows that $\min(F, \alpha)$ splits over K for each $\alpha \in K$. Thus, K is the splitting field over F of $\{\min(F, a) : a \in K\}$, so K is normal over F. □

One useful consequence of Proposition 3.28 is that if K is normal over F, then K is the splitting field of $\{\min(F, a) : a \in K\}$ by condition 4. This is perhaps the most useful criterion to show that an extension is normal.

Problems

1. Show that K is a splitting field over F for a set $\{f_1, \dots, f_n\}$ of polynomials in $F[x]$ if and only if K is a splitting field over F for the single polynomial $f_1 \cdots f_n$.

2. Let K be a splitting field of a set S of polynomials over F. If L is a subfield of K containing F for which each $f \in S$ splits over L, show that $L = K$.

3. If $F \subseteq L \subseteq K$ are fields, and if K is a splitting field of $S \subseteq F[x]$ over F, show that K is also a splitting field for S over L.

4. (a) Let K be an algebraically closed field extension of F. Show that the algebraic closure of F in K is an algebraic closure of F.

   (b) If $\mathbb{A} = \{a \in \mathbb{C} : a \text{ is algebraic over } \mathbb{Q}\}$, then, assuming that $\mathbb{C}$ is algebraically closed, show that $\mathbb{A}$ is an algebraic closure of $\mathbb{Q}$.

5. Give an example of fields $F \subseteq K \subseteq L$ where L/K and K/F are normal but L/F is not normal.

6. Let $f(x)$ be an irreducible polynomial over F of degree n, and let K be a field extension of F with $[K : F] = m$. If $\gcd(n, m) = 1$, show that f is irreducible over K.

7. Show that $x^5 - 9x^3 + 15x + 6$ is irreducible over $\mathbb{Q}(\sqrt{2}, \sqrt{3})$.

8. Find the degree of the splitting field of $x^6 + 1$ over
   (a) $\mathbb{Q}$,
   (b) $\mathbb{F}_2$.

9. Determine the splitting field of $x^4 - 7$ over
   (a) $\mathbb{Q}$,
   (b) $\mathbb{F}_5$,
   (c) $\mathbb{F}_{11}$.

10. Let F be a field, and let $f(x) \in F[x]$ be a polynomial of prime degree. Suppose for every field extension K of F that if f has a root in K, then f splits over K. Prove that either f is irreducible over F or f has a root (and hence splits) in F.

11. Show that the hypotheses of the previous problem hold for
    (a) $f(x) = x^p - a$, where $\text{char}(F) = p$ and $a \in F$.
    (b) $f(x) = x^p - x - a$, where $\text{char}(F) = p$ and $a \in F$.
    (c) $f(x) = x^p - a$, where $\text{char}(F) \neq p$ and F contains an element ω with $\omega^p = 1$ and $\omega \neq 1$.

12. Let K be a field, and suppose that $\sigma \in \text{Aut}(K)$ has infinite order. Let F be the fixed field of σ. If K/F is algebraic, show that K is normal over F.

13. Let K be a normal extension of F, and let $f(x) \in F[x]$ be an irreducible polynomial over F. Let $g_1(x)$ and $g_2(x)$ be monic irreducible factors of $f(x)$ in $K[x]$. Prove that there is a $\sigma \in \text{Gal}(K/F)$ with $\sigma(g_1) = g_2$.

14. Let K be a normal extension of F, and let $p(x)$ be an irreducible polynomial in $F[x]$. If p is not irreducible over K, show that p factors over K into a product of irreducible polynomials of the same degree. In particular, if p has a root in K, then p splits over K.

15. Let K and L be extensions of F. Show that KL is normal over F if both K and L are normal over F. Is the converse true?

16. Let M be a normal extension of F. Suppose that $a, a' \in M$ are roots of $\min(F, a)$ and that b, b' are roots of $\min(F, b)$. Determine whether or not there is an automorphism $\sigma \in \text{Gal}(M/F)$ with $\sigma(a) = a'$ and $\sigma(b) = b'$.

17. This problem will prove that any symmetric polynomial is a polynomial in the elementary symmetric functions. This problem requires some knowledge of integral ring extensions along with theorems about algebraic independence from Section 19. Let $K = k(x_1, \ldots, x_n)$ be the field of rational functions in the x_i over a field k. Then the group S_n acts as automorphisms on K as in Example 2.22. Let $f \in k[x_1, \ldots, x_n]$ be a symmetric polynomial; that is, $\sigma(f) = f$ for all $\sigma \in S_n$. Show that $f \in k[s_1, \ldots, s_n]$.
(Hint: If $F = \mathcal{F}(S_n)$, show that $F \cap k[x_1, \ldots, x_n]$ is integral over $k[s_1, \ldots, s_n]$. Moreover, show that $k[s_1, \ldots, s_n]$ is integrally closed since $k[s_1, \ldots, s_n] \cong k[x_1, \ldots, x_n]$, a fact that falls out of Section 19.)

18. Give an example of fields $k \subseteq K \subseteq L$ and $l \subseteq L$ for which l/k and L/K are algebraic, k is algebraically closed in K, and $lK = L$, but l is not algebraically closed in L.

19. This problem gives a construction of an algebraic closure of a field, due to E. Artin. Let F be a field, and let S be the set of all monic irreducible polynomials in $F[x]$. Let $A = F[x_f : f \in S]$ be a polynomial ring with one variable for each polynomial in S. Let I be the ideal of A generated by all $f(x_f)$ for $f \in S$. Show that $I \neq A$. Let $M \supseteq I$ be a maximal ideal of A, and let $F_1 = A/M$. Then F_1 is an extension of F in which each $f \in S$ has a root. Given the field F_i, construct the field F_{i+1} by repeating this procedure starting with F_i as the base field in place of F. Let $L = \bigcup_{n=1}^{\infty} F_n$. Show that each $f \in S$ splits into linear factors over L, and show that the algebraic closure of F in L is an algebraic closure of F.

4 Separable and Inseparable Extensions

Recall from Corollary 2.17 that an algebraic extension $F(a)/F$ fails to be Galois if either $\min(F, a)$ does not split over $F(a)$ or if $\min(F, a)$ has repeated roots. In the previous section, we investigated field extensions K/F for which $\min(F, a)$ splits over K for each $a \in K$. In this section, we investigate when a minimal polynomial has repeated roots. We point out that in the case of fields of characteristic 0, there is no problem of repeated roots, as we show below.

Let $f(x) \in F[x]$. A root α of f has *multiplicity* m if $(x - \alpha)^m$ divides $f(x)$ but $(x-\alpha)^{m+1}$ does not divide f. If $m > 1$, then α is called a *repeated root* of f.

Definition 4.1 *Let F be a field. An irreducible polynomial $f(x) \in F[x]$ is separable over F if f has no repeated roots in any splitting field. A polynomial $g(x) \in F[x]$ is separable over F if all irreducible factors of g are separable over F.*

Example 4.2 The polynomial $x^2 - 2$ is separable over $\mathbb{Q}$, as is $(x-1)^9$. The polynomial $x^2 + x + 1$ is separable over $\mathbb{F}_2$, since we saw in Example 2.8 that if α is a root, then so is $\alpha + 1$. Suppose that $\text{char}(F) = p$ and $a \in F - F^p$. Then $x^p - a$ is irreducible over F (see Problem 5), but it is not separable over F, since it has at most one root in any extension field of F. Note that if α is a root of $x^p - a$, then $x^p - a$ is separable over $F(\alpha)$.

The following lemma gives some basic properties of separability.

Lemma 4.3 *Let $f(x)$ and $g(x)$ be polynomials over a field F.*

1. *If f has no repeated roots in any splitting field, then f is separable over F.*
2. *If g divides f and if f is separable over F, then g is separable over F.*
3. *If $f_1, \dots, f_n$ are separable polynomials over F, then the product $f_1 \cdots f_n$ is separable over F.*
4. *If f is separable over F, then f is separable over any extension field of F.*

Proof. For property 1, if f has no repeated roots in any splitting field, then neither does any irreducible factor of f. Thus, f is separable over F. To show property 2, if g divides f with f separable over F, then no irreducible factor of f has a repeated root. However, the irreducible factors of g are also irreducible factors of f. Thus, g is separable over F. To prove property 3, we see that the set of irreducible factors of the f_i is precisely the set of irreducible factors of the polynomial $f_1 \cdots f_n$. Each of these irreducible factors have no repeated roots, so $f_1 \cdots f_n$ is separable over F. Finally, for property 4, let $f(x) \in F[x]$ be separable over F, and let K be an extension of F. If $p(x)$ is an irreducible factor of $f(x)$ in $K[x]$, let α be a root of p in some algebraic closure of K, and set $q(x) = \min(F, \alpha)$. Then $q(x) \in K[x]$, so p divides q. But q has no repeated roots, since q is an irreducible factor of f. Thus, p has no repeated roots, so f is separable over K. □

In order to have an effective test for separability, we need the concept of polynomial differentiation. A more general notion of differentiation, that of a derivation, will be used to study transcendental extensions in Chapter V.

Definition 4.4 *If $f(x) = a_0 + a_1 x + \cdots + a_n x^n \in F[x]$, then the formal derivative $f'(x)$ is defined by $f'(x) = a_1 + 2a_2 x + \cdots + n a_n x^{n-1}$.*

The formal derivative of a polynomial is well defined for any field F. We do not need limits in order to define it, as we do in calculus. However,

some strange things can happen in prime characteristic. For instance, the derivative of x^p is 0 if the base field has characteristic p.

The formal derivative satisfies the same basic properties as the derivative of calculus. If $f(x), g(x) \in F[x]$ and $a, b \in F$, then

1. $(af(x) + bg(x))' = af'(x) + bg'(x)$;
2. $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$;
3. $(f(g(x))' = f'(g(x))g'(x)$.

The proof of these properties is straightforward and is left to Problem 1.

By using derivatives, we obtain a good test for determining when a polynomial has a repeated root. This test is given in the following proposition.

Proposition 4.5 *Let $f(x) \in F[x]$ be a nonconstant polynomial. Then f has no repeated roots in a splitting field if and only if* $\gcd(f, f') = 1$ *in* $F[x]$.

Proof. We first point out that f and f' are relatively prime in $F[x]$ if and only if they are relatively prime in $K[x]$. To prove this, suppose that $\gcd(f, f') = 1$ in $F[x]$. Then there are polynomials $g, h \in F[x]$ with $1 = fg + f'h$. This also is an equation in $K[x]$, so the gcd in $K[x]$ of f and f' must divide 1. Thus, $\gcd(f, f') = 1$ in $K[x]$. Conversely, suppose that $\gcd(f, f') = 1$ in $K[x]$. If d is the gcd of f and f' in $F[x]$, then $d \in K[x]$, so d divides 1; thus, f and f' are relatively prime in $F[x]$.

Suppose that f and f' are relatively prime in $F[x]$. In particular, let K be a splitting field of $\{f, f'\}$ over F. If f and f' have a common root $\alpha \in K$, then $x - \alpha$ divides both f and f' in $K[x]$. This would contradict the fact that f and f' are relatively prime in $K[x]$. Therefore, f and f' have no common roots.

Conversely, if f and f' have no common roots in a splitting field K of $\{f, f'\}$, let $d(x)$ be the greatest common divisor in $K[x]$ of $f(x)$ and $f'(x)$. Then d splits over K since f splits over K and d divides f. Any root of d is then a common root of f and f' since d also divides f'. Thus, $d(x)$ has no roots, so $d = 1$. Therefore, f and f' are relatively prime over K; hence, they are also relatively prime over F. □

With this derivative test, we can give the following criteria for when a polynomial is separable. Note that this test does not require that we know the roots of a polynomial.

Proposition 4.6 *Let $f(x) \in F[x]$ be an irreducible polynomial.*

1. *If* $\mathrm{char}(F) = 0$*, then f is separable over F. If* $\mathrm{char}(F) = p > 0$*, then f is separable over F if and only if $f'(x) \neq 0$, and this occurs if and only if $f(x) \notin F[x^p]$.*

2. *If* $\text{char}(F) = p$, *then* $f(x) = g(x^{p^m})$ *for some integer* $m \geq 0$ *and some* $g(x) \in F[x]$ *that is irreducible and separable over* F.

Proof. If $f(x) \in F[x]$ is irreducible over F, then the only possibility for $\gcd(f, f')$ is 1 or f. If $\text{char}(F) = 0$, then $\deg(f') = \deg(f) - 1$; thus, f does not divide f', and so $\gcd(f, f') = 1$. Therefore, by Proposition 4.5, f has no repeated roots, so f is separable over F. If $\text{char}(F) = p > 0$, the same reasoning shows $\gcd(f, f') = f$ if and only if f divides f', if and only if $f'(x) = 0$, if and only if $f(x) \in F[x^p]$.

For statement 2, suppose that $\text{char}(F) = p$, and let $f(x) \in F[x]$. Let m be maximal such that $f(x) \in F[x^{p^m}]$. Such an m exists, since $f \in F[x^{p^0}]$ and f lies in $F[x^{p^r}]$ for only finitely many r because any nonconstant polynomial in $F[x^{p^r}]$ has degree at least p^r. Say $f(x) = g(x^{p^m})$. Then $g(x) \notin F[x^p]$ by maximality of m. Moreover, $g(x)$ is irreducible over F, since if $g(x) = h(x) \cdot k(x)$, then $f(x) = h(x^{p^m}) \cdot k(x^{p^m})$ is reducible over F. By statement 2, g is separable over F. □

We now extend the concept of separability to field elements and field extensions.

Definition 4.7 *Let* K *be an extension field of* F *and let* $\alpha \in K$. *Then* α *is separable over* F *if* $\min(F, \alpha)$ *is separable over* F. *If every* $\alpha \in K$ *is separable over* F, *then* K *is separable over* F.

Example 4.8 If F is a field of characteristic 0, then any algebraic extension of F is separable over F, since every polynomial in $F[x]$ is separable over F. If k is a field of characteristic $p > 0$ and if $k(x)$ is the rational function field in one variable over k, then the extension $k(x)/k(x^p)$ is not separable, for $\min(k(x^p), x) = t^p - x^p$, which has only x as a root.

We are now in a position to give a characterization of Galois extension. This characterization is the most common way to show that a field extension is Galois.

Theorem 4.9 *Let* K *be an algebraic extension of* F. *Then the following statements are equivalent:*

1. K *is Galois over* F.
2. K *is normal and separable over* F.
3. K *is a splitting field of a set of separable polynomials over* F.

Proof. (1) $\Rightarrow$ (2): Suppose that K is Galois over F, and let $\alpha \in K$. Let $\alpha_1, \ldots, \alpha_n$ be the distinct elements of the set $\{\sigma(\alpha) : \sigma \in \text{Gal}(K/F)\}$. This set is finite by Lemma 2.3, since each $\sigma(\alpha)$ is a root of $\min(F, \alpha)$. Let

$f(x) = \prod_i (x - \alpha_i) \in K[x]$. Then $\tau(f) = f$, since τ permutes the α_i. Thus, the coefficients of f lie in $\mathcal{F}(\text{Gal}(K/F)) = F$, so $f(x) \in F[x]$. Therefore, $\min(F, \alpha)$ divides f, and so $\min(F, \alpha)$ splits over K and has no repeated roots. Since this is true for each $\alpha \in K$, the field K is the splitting field of the set $\{\min(F, \alpha) : \alpha \in K\}$ of polynomials separable over F. Hence, K/F is normal and separable.

(2) $\Rightarrow$ (3): If K/F is normal and separable, then K is the splitting field of the set of separable polynomials $\{\min(F, \alpha) : \alpha \in K\}$ by Proposition 3.28.

(3) $\Rightarrow$ (1): We first assume that $[K : F] < \infty$, and we use induction on $n = [K : F]$. If $n = 1$, then $K = F$ is trivially Galois over F. So, suppose that $n > 1$ and that the result holds for field extensions of degree less than n. Say K is the splitting field of the set of separable polynomials $\{f_i(x)\}$. Since $n > 1$, there is a root α of one of the f_i which is not in F. Let $L = F(\alpha)$. Then $[L : F] > 1$, so $[K : L] < n$. Since K is the splitting field over L of the $\{f_i\}$, which are separable over L, by induction K is Galois over L. Let $H = \text{Gal}(K/L)$, a subgroup of $\text{Gal}(K/F)$. Let $\alpha_1, \dots, \alpha_r$ be the distinct roots of $\min(F, \alpha)$. Then, since α is separable over F, we have $[L : F] = r$. By the isomorphism extension theorem, there are $\tau_i \in \text{Gal}(K/F)$ with $\tau_i(\alpha) = \alpha_i$. The cosets $\tau_i H$ are then distinct, since if $\tau_i^{-1}\tau_j \in H = \text{Gal}(K/L)$, then $(\tau_i^{-1}\tau_j)(\alpha) = \alpha$; hence, $\alpha_i = \tau_i(\alpha) = \tau_j(\alpha) = \alpha_j$. Let $G = \text{Gal}(K/F)$. We have

$$|G| = |G : H| \cdot |H| \geq r \cdot |H| = [L : F] \cdot [K : L] = [K : F].$$

Since $|G| \leq [K : F]$ by Proposition 2.13, we get $|G| = [K : F]$, so K is Galois over F.

Now suppose that K/F is arbitrary. By hypothesis, K is the splitting field over F of a set S of separable polynomials over F. Let X be the set of roots of all of these polynomials. So, $K = F(X)$. Let $a \in \mathcal{F}(\text{Gal}(K/F))$. We wish to show that $a \in F$. There is a finite subset $\{\alpha_1, \dots, \alpha_n\} \subseteq X$ with $a \in F(\alpha_1, \dots, \alpha_n)$. Let $L \subseteq K$ be the splitting field of $\{\min(F, \alpha_i) : 1 \leq i \leq n\}$. Then, by the previous paragraph, L/F is a finite Galois extension. Note that $a \in L$. An application of the isomorphism extension theorem shows that each element of $\text{Gal}(L/F)$ extends to an F-automorphism of K, and so Proposition 3.28 implies that

$$\text{Gal}(L/F) = \{\sigma|_L : \sigma \in \text{Gal}(K/F)\}.$$

Therefore, $a \in \mathcal{F}(\text{Gal}(L/F))$, and this fixed field is F, since L/F is Galois. This proves $\mathcal{F}(\text{Gal}(K/F)) = F$, so K/F is Galois. □

Corollary 4.10 *Let L be a finite extension of F.*

1. *L is separable over F if and only if L is contained in a Galois extension of F.*

2. *If $L = F(\alpha_1, \dots, \alpha_n)$ with each α_i separable over F, then L is separable over F.*

Proof. If $L \subseteq K$ with K/F Galois, then K/F is separable by Theorem 4.9. Hence, L/F is separable. Conversely, suppose that L/F is separable. Since $[L : F] < \infty$, we may write $L = F(\alpha_1, \dots, \alpha_n)$, and each α_i is separable over F. If K is the splitting field of $\{\min(F, \alpha_i) : 1 \le i \le n\}$, then $L \subseteq K$, and K/F is Galois by Theorem 4.9.

For the proof of statement 2, let $L = F(\alpha_1, \dots, \alpha_n)$ with each α_i separable over F. Then each $\min(F, \alpha_i)$ is a separable polynomial over F. If K is the splitting field of these polynomials, then K/F is Galois by Theorem 4.9. Thus, again by that theorem, K is separable over F. Since $L \subseteq K$, we see that L is separable over F. □

Fields for which all algebraic extensions are separable are particularly well behaved. We now determine which fields have this property.

Definition 4.11 *A field F is perfect if every algebraic extension of F is separable.*

Example 4.12 Any field of characteristic 0 is perfect. Therefore, any field containing $\mathbb{Q}$ or contained in $\mathbb{C}$ is perfect. Any algebraically closed field is perfect for the trivial reason that there are no proper algebraic extensions of an algebraically closed field.

The following theorem characterizes perfect fields of prime characteristic. We have seen in previous examples that if $a \in F - F^p$, then $x^p - a$ is an irreducible polynomial that is not separable. Therefore, for F to be perfect, we must have $F^p = F$. We now show this is sufficient to ensure that F is perfect.

Theorem 4.13 *Let F be a field of characteristic p. Then F is perfect if and only if $F^p = F$.*

Proof. Suppose that F is perfect. Let $a \in F$, and consider the field $K = F(\alpha)$, where α is a root of $x^p - a$. The minimal polynomial of α divides $x^p - a = (x - \alpha)^p$. However, K is separable over F since F is perfect; thus, this minimal polynomial has no repeated roots. This means $\alpha \in F$, so $a \in F^p$.

Conversely, suppose that $F^p = F$. Let K be an algebraic extension of F, and let $\alpha \in K$. If $p(x) = \min(F, \alpha)$, then by Proposition 4.6 there is an m with $p(x) = g(x^{p^m})$ for some $g(x) \in F[x]$ with g irreducible and separable over F. If $g(x) = a_0 + a_1 x + \cdots + x^r$, then there are $b_i \in F$ with $b_i^p = a_i$ for all i. Then $p(x) = \sum_i b_i x^{p^m i} = (\sum_i b_i x^{p^{m-1} i})^p$. This contradicts the irreducibility of p unless $m = 1$. Thus, $p = g$ is separable over F, so α is separable over F. Therefore, any algebraic extension of F is separable, so F is perfect. □

Example 4.14 Any finite field is perfect; to prove this, let F be a finite field. The map $\varphi : F \to F$ given by $\varphi(a) = a^p$ is a nonzero field homomorphism, so φ is injective. Since F is finite, φ is also surjective. Thus, $F^p = \text{im}(\varphi) = F$, so F is perfect by Theorem 4.13. We give another proof of this fact in Corollary 6.13.

Purely inseparable extensions

We now discuss the condition diametrically opposed to separability. This situation is only relevant in prime characteristic, since any algebraic extension in characteristic 0 is separable. If F is a field of characteristic $p > 0$, and if $a \in F$, then $x^p - a$ has only one distinct root in any splitting field, since if α is a root of f, then $x^p - a = (x - \alpha)^p$. In this case, $\alpha^p = a \in F$.

Definition 4.15 *Let K be an algebraic field extension of F. An element $\alpha \in K$ is purely inseparable over F if* $\min(F, \alpha)$ *has only one distinct root. The field K is purely inseparable over F if every element in K is purely inseparable over F.*

The definition of purely inseparable requires that we know how many roots there are of a minimal polynomial of an element. The following lemma gives an easier way to determine when an element is purely inseparable over a field.

Lemma 4.16 *Let F be a field of characteristic $p > 0$. If α is algebraic over F, then α is purely inseparable over F if and only if $\alpha^{p^n} \in F$ for some n. When this happens,* $\min(F, \alpha) = (x - \alpha)^{p^n}$ *for some n.*

Proof. If $\alpha^{p^n} = a \in F$, then α is a root of the polynomial $x^{p^n} - a$. This polynomial factors over $F(\alpha)$ as $(x - \alpha)^{p^n}$, and $\min(F, \alpha)$ divides this polynomial, so $\min(F, \alpha)$ has only α as a root. Conversely, suppose that α is purely inseparable over F, and let $f(x) = \min(F, \alpha)$. There is a separable irreducible polynomial $g(x)$ over F with $f(x) = g(x^{p^m})$ by Proposition 4.6. If g factors over a splitting field as $g(x) = (x - b_1) \cdots (x - b_r)$, then $f(x) = (x^{p^m} - b_i) \cdots (x^{p^m} - b_r)$. If $r > 1$, then separability of g says that the b_i are distinct. By assumption, the only root of f is α. Thus, $b_i = \alpha^{p^m}$ for each i. Hence, $r = 1$, so $f(x) = x^{p^m} - b_1$. Therefore, $\alpha^{p^m} \in F$, and $\min(F, \alpha) = x^{p^m} - b_1 = (x - \alpha)^{p^m}$. □

The basic properties of purely inseparable extensions are given in the following lemma.

Lemma 4.17 *Let K be an algebraic extension of F.*

1. *If $\alpha \in K$ is separable and purely inseparable over F, then $\alpha \in F$.*

2. *If K/F is purely inseparable, then K/F is normal and* $\text{Gal}(K/F) = \{\text{id}\}$. *Moreover, if $[K : F] < \infty$, and if $p = \text{char}(F)$, then $[K : F] = p^n$ for some n.*

3. *If $K = F(X)$ with each $\alpha \in X$ purely inseparable over F, then K is purely inseparable over F.*

4. *If $F \subseteq L \subseteq K$ are fields, then K/F is purely inseparable if and only if K/L and L/F are purely inseparable.*

Proof. Suppose that $\alpha \in K$ is both separable and purely inseparable over F. Then $\min(F, \alpha)$ has only one distinct root, and it also has no repeated roots. Therefore, $p(x) = x - \alpha$, so $\alpha \in F$.

For property 2, if K/F is purely inseparable, then each $\min(F, \alpha)$ splits over K, since the only root of $\min(F, \alpha)$ is α itself. Consequently, K is normal over F by Proposition 3.28. If $\sigma \in \text{Gal}(K/F)$, then, for any $\alpha \in K$, the automorphism σ maps α to a root of $\min(F, \alpha)$. Thus, $\sigma(\alpha) = \alpha$, so $\sigma = \text{id}$. Therefore, $\text{Gal}(K/F) = \{\text{id}\}$. If $[K : F] < \infty$, then K is finitely generated over F; say, $K = F(\alpha_1, \dots, \alpha_n)$. To prove that $[K : F]$ is a power of $p = \text{char}(F)$, by Proposition 1.20 it suffices by induction to prove this in the case $K = F(\alpha)$. But then $[K : F] = \deg(\min(F, \alpha))$, which is a power of p by the previous lemma.

To prove property 3, suppose that K is generated over F by a set X of elements purely inseparable over F. Let $a \in K$. Then $a \in F(\alpha_1, \dots, \alpha_n)$ for some $\alpha_i \in X$. Since each α_i is purely inseparable over F, there is an m such that $\alpha_i^{p^m} \in F$ for each i. Because a is a polynomial in the α_i, we see that $a^{p^m} \in F$. This forces $\min(F, a)$ to divide $(x - a)^{p^m}$; hence, $\min(F, a)$ has only one distinct root. Therefore, a is purely inseparable over F, and so K/F is purely inseparable.

Finally, for property 4, if K/F is purely inseparable, then for any $a \in K$, there is an m with $a^{p^m} \in F$. Thus, $a^{p^m} \in L$, so K/L is purely inseparable. It is clear that L/F is purely inseparable. Conversely, if L/F and K/L are purely inseparable, let $a \in K$. Then $a^{p^m} \in L$ for some m, and so $(a^{p^m})^{p^r} = a^{p^{m+r}} \in F$ for some r. Therefore, K/F is purely inseparable. □

Example 4.18 A field extension need not be either separable or purely inseparable. For instance, if $F = \mathbb{F}_2(x)$ is the rational function field in one variable over $\mathbb{F}_2$, and if $K = F(x^{1/6})$, then $K = F(\sqrt{x}, \sqrt[3]{x})$. Moreover, $\sqrt{x}$ is purely inseparable over F, and $\sqrt[3]{x}$ is separable over F. The subfield $F(\sqrt{x})$ is purely inseparable over F, and the subfield $F(\sqrt[3]{x})$ is separable over F.

In the previous example, we can show that $F(\sqrt[3]{x})$ consists of all the elements of K that are separable over F and that $F(\sqrt{x})$ consists of all the elements of K that are purely inseparable over F. This is a special case of the following lemma. We first give the relevant definitions.

Definition 4.19 *Let K be a field extension of F. Then the separable closure of F in K is the set $\{a \in K : a \text{ is separable over } F\}$. The purely inseparable closure of F in K is the set $\{a \in K : a \text{ is purely inseparable over } F\}$.*

The separable and purely inseparable closures of F in K are fields, as we now show.

Proposition 4.20 *Let K be a field extension of F. If S and I are the separable and purely inseparable closures of F in K, respectively, then S and I are field extensions of F with S/F separable, I/F purely inseparable, and $S \cap I = F$. If K/F is algebraic, then K/S is purely inseparable.*

Proof. Let $a, b \in S$. Then $F(a, b)$ is a separable extension of F by Lemma 4.10. Hence, $a \pm b$, ab, and a/b are separable over F, so they all lie in S. Thus, S is a field. For I, if $c, d \in I$, then there are n, m with $c^{p^n} \in F$ and $d^{p^m} \in F$. Setting $N = nm$, we have $(c \pm d)^{p^N}$, $(cd)^{p^N}$, and $(c/d)^{p^N} \in F$. Thus, $c \pm d$, cd, and c/d belong to I, so I is a field. The equality $S \cap I = F$ holds, since $S \cap I$ is both separable and purely inseparable over F. Finally, suppose that K/F is algebraic. If $\alpha \in K$, then $\min(F, \alpha) = g(x^{p^n})$ for some separable, irreducible polynomial $g(x) \in F[x]$ by Proposition 4.6. If $a = \alpha^{p^n}$, then $g(a) = 0$, so $g(x) = \min(F, a)$. Therefore, a is separable over F, so $\alpha^{p^n} = a \in S$. Thus, K/S is purely inseparable. □

If K/F is an algebraic extension, we can break up the extension K/F into a separable extension S/F followed by a purely inseparable extension K/S, where S is the separable closure of F in K. Use of the separable closure is a nice tool to prove results dealing with separability. As an illustration, we prove that separability is a transitive property.

Proposition 4.21 *If $F \subseteq L \subseteq K$ are fields such that L/F and K/L are separable, then K/F is separable.*

Proof. Let S be the separable closure of F in K. Then $L \subseteq S$, as L/F is separable. Also, since K/L is separable, K/S is separable. But K/S is purely inseparable, so $K = S$. Thus, K is separable over F. □

Example 4.22 Let K be a finite extension of F, and suppose that $\text{char}(F)$ does not divide $[K : F]$. We show that K/F is separable. If $\text{char}(F) = 0$, then this is clear, so suppose that $\text{char}(F) = p > 0$. Let S be the separable closure of F in K. Then K/S is purely inseparable, so $[K : S] = p^n$ for some n by Lemma 4.17. However, since p does not divide $[K : F]$, this forces $[K : S] = 1$. Thus, $K = S$, so K is separable over F.

A natural question that Proposition 4.20 raises is whether the extension K/I is separable. The answer in general is no, although it is true if K/F is normal, as we now show.

Theorem 4.23 *Let K be a normal extension of F, and let S and I be the separable and purely inseparable closures of F in K, respectively. Then S/F is Galois, $I = \mathcal{F}(\text{Gal}(K/F))$, and $\text{Gal}(S/F) \cong \text{Gal}(K/I)$. Thus, K/I is Galois. Moreover, $K = SI$.*

Proof. Let $a \in S$, and set $f(x) = \min(F, a)$. Since K is normal over F, the polynomial f splits over K. Since a is separable over F, the polynomial f has no repeated roots, so all its roots are separable over S. Thus, f splits over S. Hence, S is normal over F by Proposition 3.28, and since S is separable over F, we see by Theorem 4.9 that S is Galois over F. The map $\theta : \text{Gal}(K/F) \longrightarrow \text{Gal}(S/F)$ given by $\theta(\sigma) = \sigma|_S$ is a well-defined group homomorphism. The kernel of θ is $\text{Gal}(K/S)$, and this group is trivial by Lemma 4.17 since K is purely inseparable over S. By the isomorphism extension theorem, if $\tau \in \text{Gal}(S/F)$, there is a $\sigma \in \text{Gal}(K/F)$ with $\sigma|_S = \tau$. Thus, θ is an isomorphism.

To show that $I = \mathcal{F}(\text{Gal}(K/F))$, if $a \in I$, then $a^{p^n} \in F$ for some n. For $\sigma \in \text{Gal}(K/F)$, we have $a^{p^n} = \sigma(a^{p^n}) = \sigma(a)^{p^n}$, so $\sigma(a) = a$. Thus, $I \subseteq \mathcal{F}(\text{Gal}(K/F))$. Conversely, take $b \in \mathcal{F}(\text{Gal}(K/F))$. There is an n with $b^{p^n} \in S$ because K/S is purely inseparable. Let $\tau \in \text{Gal}(S/F)$. Since θ is surjective, there is a $\sigma \in \text{Gal}(K/F)$ with $\tau = \theta(\sigma) = \sigma|_S$. Then $\tau(b^{p^n}) = \sigma(b^{p^n}) = b^{p^n}$. This is true for each τ; hence, $b^{p^n} \in \mathcal{F}(\text{Gal}(S/F)) = F$. This equality holds since S is Galois over F. Thus, b is purely inseparable over F. This proves $I = \mathcal{F}(\text{Gal}(K/F))$, so $\text{Gal}(K/F) = \text{Gal}(K/I)$. Therefore, K is Galois over I; hence, K/I is separable. Finally, K is separable over SI since $I \subseteq SI$, and K is purely inseparable over SI since $S \subseteq SI$. Therefore, $K = SI$. □

Let K be a finite extension of F. If S and I are the separable and purely inseparable closures of F in K, respectively, we define the *separable degree* $[K : F]_s$ of K/F to be $[S : F]$ and the *inseparable degree* $[K : F]_i$ to be $[K : S]$. With these definitions, we see that $[K : F]_s[K : F]_i = [K : F]$. By Theorem 4.23, if K/F is normal, then $[K : I] = [S : F]$, and so $[K : S] = [I : F]$. However, as the example below shows, in general $[K : S] \neq [I : F]$. The inseparable degree is defined to be $[K : S]$ and not $[I : F]$ because the degree $[K : S]$ is a better measure for how far the extension K/F is from being separable. The example below shows that it is possible to have $I = F$ even if K is not separable over F. We will use the concepts of separable and inseparable degrees in Section 8.

Example 4.24 We give an example of a field extension K/F in which K is not separable over the purely inseparable closure I of F in K. This is also an example of a nonseparable field extension K/F in which the purely inseparable closure is F. Let k be a field of characteristic 2, let F be the rational function field $F = k(x, y)$, let $S = F(u)$, where u is a root of t^2+t+x, and let $K = S(\sqrt{uy})$. Then K/S is purely inseparable and S/F is

separable, so S is the separable closure of F in K. We will show that $I = F$, which will prove that K/I is not separable since K/S is not separable. To do this, we show that if $a \in K$ with $a^2 \in F$, then $a \in F$. A basis for K/F is 1, u, $\sqrt{uy}$, and $u\sqrt{uy}$. Say $a^2 \in F$ and write $a = \alpha + \beta u + \gamma\sqrt{uy} + \delta u\sqrt{uy}$ with $\alpha, \beta, \gamma, \delta \in F$. Then

$$a^2 = \alpha^2 + \beta^2(u + x) + \gamma^2(uy) + \delta^2(u + x)uy.$$

The coefficient of u is zero since $a^2 \in F$, so

$$\beta^2 + (\gamma + \delta)^2 y + \delta^2 xy = 0.$$

If $\delta = 0$, then $\beta^2 + \gamma^2 y = 0$, so $\gamma = 0$ since y is not a square in F. But then $\beta = 0$, so $a \in F$. If $\delta \neq 0$, then

$$x = \frac{\beta^2 + (\gamma + \delta)^2 y}{\delta^2 y} = \left(\frac{\gamma}{\delta} + 1\right)^2 + \left(\frac{\beta}{\delta}\right)^2 y,$$

which means that $x \in F^2(y)$. But this is impossible. Thus, $\delta = 0$, and so we conclude that $a \in F$. Thus, $I = F$, so K/I is not separable. Note that $K \neq SI$ also.

Problems

1. Prove the sum, product, and chain rules for formal polynomial differentiation in $F[x]$.

2. If $F \subseteq L \subseteq K$ are fields such that K/F is separable, show that L/F and K/L are separable.

3. If K is a field extension of F and if $\alpha \in K$ is not separable over F, show that α^{p^m} is separable over F for some $m \geq 0$, where $p = \text{char}(F)$.

4. Let $F \subseteq L \subseteq K$ be fields such that K/L is normal and L/F is purely inseparable. Show that K/F is normal.

5. Let F be a field of characteristic $p > 0$, and let $a \in F - F^p$. Show that $x^p - a$ is irreducible over F.

6. Let F be a field of characteristic $p > 0$, and let K be a purely inseparable extension of F with $[K : F] = p^n$. Prove that $a^{p^n} \in F$ for all $a \in K$.

7. Let K and L be extensions of F. Show that KL is separable over F if both K and L are separable over F. Is the converse true?

8. Let K and L be extensions of F. Show that KL is purely inseparable over F if both K and L are purely inseparable over F. Is the converse true?

9. Let K and L be extensions of F. Show that KL is Galois over F if both K and L are Galois over F. Is the converse true?

10. Let K and L be subfields of a common field, both of which contain a field F. Prove the following statements.

    (a) If $K = F(X)$ for some set $X \subseteq K$, then $KL = L(X)$.

    (b) $[KL : F] \leq [K : F] \cdot [L : F]$.

    (c) If K and L are algebraic over F, then KL is algebraic over F.

    (d) Prove that the previous statement remains true when "algebraic" is replaced by "normal," "separable," "purely inseparable," or "Galois."

11. Let K be the rational function field $k(x)$ over a perfect field k of characteristic $p > 0$. Let $F = k(u)$ for some $u \in K$, and write $u = f(x)/g(x)$ with f and g relatively prime. Show that K/F is a separable extension if and only if $u \notin K^p$.

12. Let K be a finite extension of F with $\operatorname{char} F = p > 0$ and $K^p \subseteq F$. Thus, K/F is purely inseparable. A set $\{a_1, \ldots, a_n\} \subseteq K$ is said to be a *p-basis* for K/F provided that there is a chain of proper extensions

$$F \subset F(a_1) \subset \cdots \subset F(a_n) = K.$$

Show that if $\{a_1, \ldots, a_n\}$ is a p-basis for K/F, then $[K : F] = p^n$, and conclude that the number of elements in a p-basis is uniquely determined by K/F. The number n is called the *p-dimension* of K/F. Also, show that any finite purely inseparable extension has a p-basis.

13. Give three examples of a field extension K/F which is neither normal nor separable. Note that two such examples are given in the section.

14. Let k be a field of characteristic $p > 0$, let $K = k(x, y)$ be the rational function field over k in two variables, and let $F = k(x^p, y^p)$. Show that K/F is a purely inseparable extension of degree p^2. Show that $K \neq F(a)$ for any $a \in K$.

15. Prove the following product formulas for separability and inseparability degree: If $F \subseteq L \subseteq K$ are fields, then show that $[K : F]_s = [K : L]_s[L : F]_s$ and $[K : F]_i = [K : L]_i[L : F]_i$.

5 The Fundamental Theorem of Galois Theory

We are now in the position to prove the fundamental theorem of Galois theory, which describes the intermediate fields of a Galois extension K/F in terms of the subgroups of the Galois group $\mathrm{Gal}(K/F)$. This theorem allows us to translate many questions about fields into questions about finite groups. As an application of this theorem, we give a mostly algebraic proof of the fundamental theorem of algebra, which says that the complex field $\mathbb{C}$ is algebraically closed.

Theorem 5.1 (Fundamental Theorem of Galois Theory) *Let K be a finite Galois extension of F, and let $G = \mathrm{Gal}(K/F)$. Then there is a 1–1 inclusion reversing correspondence between intermediate fields of K/F and subgroups of G, given by $L \mapsto \mathrm{Gal}(K/L)$ and $H \mapsto \mathcal{F}(H)$. Furthermore, if $L \leftrightarrow H$, then $[K : L] = |H|$ and $[L : F] = [G : H]$. Moreover, H is normal in G if and only if L is Galois over F. When this occurs, $\mathrm{Gal}(L/F) \cong G/H$.*

Proof. We have seen in Lemma 2.9 that the maps $L \mapsto \mathrm{Gal}(K/L)$ and $H \mapsto \mathcal{F}(H)$ give injective inclusion reversing correspondences between the set of fixed fields L with $F \subseteq L \subseteq K$ and the set of subgroups of G of the form $\mathrm{Gal}(K/L)$ for some L with $F \subseteq L \subseteq K$. Let L be a subfield of K containing F. Since K is Galois over F, the extension K is normal and separable over F. Thus, K is also normal and separable over L, so K is Galois over L. Hence, $L = \mathcal{F}(\mathrm{Gal}(K/L))$, so any intermediate field is a fixed field. Also, if H is a subgroup of G, then H is a finite group, so $H = \mathrm{Gal}(K/\mathcal{F}(H))$ by Proposition 2.14. Every subgroup of G is therefore such a Galois group. The maps above then yield the desired correspondences. Recall that $|\mathrm{Gal}(K/F)| = [K : F]$ if K is Galois over F by Proposition 2.14. Thus, if $L \leftrightarrow H$, we have $|H| = [K : L]$, since K is Galois over L and $H = \mathrm{Gal}(K/L)$. Therefore,

$$[G : H] = |G|/|H| = [K : F]/[K : L] = [L : F].$$

Suppose that H is normal in G, and let $L = \mathcal{F}(H)$. Take $a \in L$, and let b be any root of $\min(F, a)$ in K. By the isomorphism extension theorem, there is a $\sigma \in G$ with $\sigma(a) = b$. If $\tau \in H$, then $\tau(b) = \sigma(\sigma^{-1}\tau\sigma(a))$. However, since H is normal in G, the element $\sigma^{-1}\tau\sigma \in H$, so $\sigma^{-1}\tau\sigma(a) = a$. Thus, $\tau(b) = \sigma(a) = b$, so $b \in \mathcal{F}(H) = L$. Since $\min(F, a)$ splits over K, this shows that $\min(F, a)$ actually splits over L. Therefore, L is normal over F by Proposition 3.28. Since K/F is separable and $L \subseteq K$, the extension L/F is also separable, and so L is Galois over F. Conversely, suppose that L is Galois over F. Let $\theta : G \longrightarrow \mathrm{Gal}(L/F)$ be given by $\theta(\sigma) = \sigma|_L$. Normality of L/F shows that $\sigma|_L \in \mathrm{Gal}(L/F)$ by Proposition 3.28, so θ is a well-defined group homomorphism. The kernel of θ is

$$\ker(\theta) = \{\sigma \in K : \sigma|_L = \mathrm{id}\} = \mathrm{Gal}(K/L) = H.$$

Therefore, H is normal in G. The map θ is surjective since, if $\tau \in \text{Gal}(L/F)$, then there is a $\sigma \in G$ with $\sigma|_L = \tau$ by the isomorphism extension theorem. Thus, $\text{Gal}(L/F) \cong G/H$. □

Given a Galois extension K/F, on the surface it would seem to be intractable to determine all intermediate fields; the main problem is knowing whether we have found all of them. However, the Galois group $G = \text{Gal}(K/F)$ is a finite group, which means that there is a systematic way of finding all subgroups of G. By finding all subgroups, we can then determine the fixed fields of each, thereby having all intermediate fields by the fundamental theorem. The next two examples illustrate this procedure. Of course, if G is large, it may be too complicated to find all subgroups of G.

Example 5.2 The field $\mathbb{Q}(\sqrt[3]{2}, \omega)$ is Galois over $\mathbb{Q}$, as we have seen previously. The Galois group is a group of order 6. From group theory, there are two nonisomorphic groups of order 6: the cyclic group $\mathbb{Z}/6\mathbb{Z}$ and the symmetric group S_3. Which is the Galois group? The subfield $\mathbb{Q}(\sqrt[3]{2})$ is not Galois over $\mathbb{Q}$, since the minimal polynomial of $\sqrt[3]{2}$ does not split over $\mathbb{Q}(\sqrt[3]{2})$. Therefore, the corresponding subgroup is not normal in G. However, every subgroup of an Abelian group is normal, so our Galois group is non-Abelian. Thus, $G = \text{Gal}(\mathbb{Q}(\sqrt[3]{2}, \omega)/\mathbb{Q}) \cong S_3$. We can also explicitly demonstrate this isomorphism. By the isomorphism extension theorem, there are $\mathbb{Q}$-automorphisms σ, τ of $\mathbb{Q}(\sqrt[3]{2}, \omega)$ with

$$\begin{aligned} \sigma &: \sqrt[3]{2} \to \omega\sqrt[3]{2}, \quad \omega \to \omega, \\ \tau &: \sqrt[3]{2} \to \sqrt[3]{2}, \quad \omega \to \omega^2. \end{aligned}$$

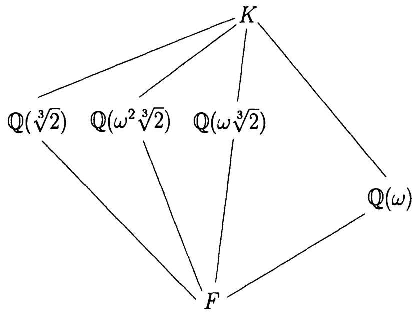
It is easy to check that σ has order 3, τ has order 2, and $\sigma\tau \neq \tau\sigma$. The subgroups of the Galois group are then

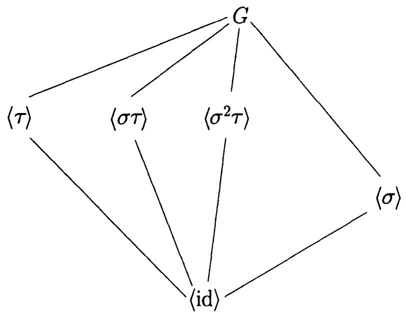
$$\langle \text{id} \rangle, \langle \sigma \rangle, \langle \tau \rangle, \langle \sigma\tau \rangle, \langle \sigma^2\tau \rangle, G.$$

The corresponding fixed fields are

$$\mathbb{Q}(\sqrt[3]{2}, \omega),\ \mathbb{Q}(\omega),\ \mathbb{Q}(\sqrt[3]{2}),\ \mathbb{Q}(\omega^2\sqrt[3]{2}),\ \mathbb{Q}(\omega\sqrt[3]{2}),\ \mathbb{Q}.$$

One way to verify that these fields are in fact the correct ones is to show that, for any of these fields, the field is indeed fixed by the appropriate subgroup and its dimension over $\mathbb{Q}$ is correct. For instance, $\sqrt[3]{2}$ is fixed by τ; hence, $\mathbb{Q}(\sqrt[3]{2}) \subseteq \mathcal{F}(\tau)$. Since the index $[G : \langle \tau \rangle] = 3$, we must have $[\mathcal{F}(\tau) : F] = 3$. But $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$, so $\mathbb{Q}(\sqrt[3]{2}) = \mathcal{F}(\tau)$. This use of dimension is extremely useful in determining the fixed field of a subgroup.





Example 5.3 Let $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. Then K is the splitting field of $\{x^2 - 2, x^2 - 3\}$ over $\mathbb{Q}$ or, alternatively, the splitting field of $(x^2-2)(x^2-3)$ over $\mathbb{Q}$. The dimension of $K/\mathbb{Q}$ is 4. The four automorphisms of $K/\mathbb{Q}$ are given by

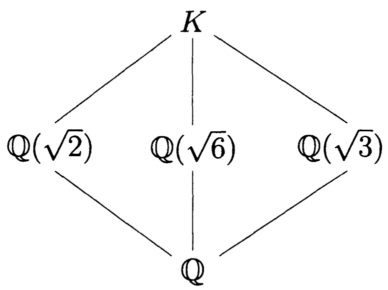
$$\begin{aligned}
\mathrm{id} : \sqrt{2} \to \sqrt{2}, &\quad \sqrt{3} \to \sqrt{3}, \\
\sigma : \sqrt{2} \to -\sqrt{2}, &\quad \sqrt{3} \to \sqrt{3}, \\
\tau : \sqrt{2} \to \sqrt{2}, &\quad \sqrt{3} \to -\sqrt{3}, \\
\sigma\tau : \sqrt{2} \to -\sqrt{2}, &\quad \sqrt{3} \to -\sqrt{3}.
\end{aligned}$$

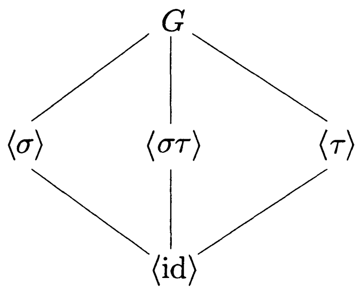
This Galois group is Abelian and is isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. The subgroups of $G = \mathrm{Gal}(K/\mathbb{Q})$ are

$$\langle \mathrm{id} \rangle, \langle \sigma \rangle, \langle \tau \rangle, \langle \sigma\tau \rangle, G.$$

The corresponding intermediate fields are

$$K,\ \mathbb{Q}(\sqrt{3}),\ \mathbb{Q}(\sqrt{2}),\ \mathbb{Q}(\sqrt{6}),\ \mathbb{Q}.$$





Example 5.4 Let $F = \mathbb{C}(t)$ be the rational function field in one variable over $\mathbb{C}$, and let $f(x) = x^n - t \in F[x]$. The polynomial f is irreducible over F by the Eisenstein criterion, since F is the quotient field of the unique

factorization domain $\mathbb{C}[t]$ and t is an irreducible element of $\mathbb{C}[t]$. Let K be the splitting field of f over F. Then $K = F(\alpha)$, where α is any root of $f(x)$. To see this, if $\omega = \exp(2\pi i/n)$, then $\omega^n = 1$, so $\omega^i\alpha$ is a root of $f(x)$ for each i. There are exactly n distinct powers of ω, so the n distinct elements $\alpha, \omega\alpha, \dots, \omega^{n-1}\alpha$ are precisely the roots of f. All of these lie in $F(\alpha)$ and generate $F(\alpha)$, so $K = F(\alpha)$. The extension K/F is then Galois since f has no repeated roots. We see that $[K : F] = \deg(f) = n$.

The isomorphism extension theorem tells us that there is an automorphism σ of K defined by $\sigma(\alpha) = \omega\alpha$. This formula yields that $\sigma^i(\alpha) = \omega^i\alpha$ for each i, so $\sigma^i(\alpha) = \alpha$ if and only if n divides i. Thus, σ has order n in $\mathrm{Gal}(K/F)$. This forces $\mathrm{Gal}(K/F)$ to be the cyclic group generated by σ. Each subgroup of $\langle\sigma\rangle$ is cyclic and can be generated by an element σ^m with m a divisor of n. Given a divisor m of n, if $n = mk$, then the element α^k is fixed by σ^m, since

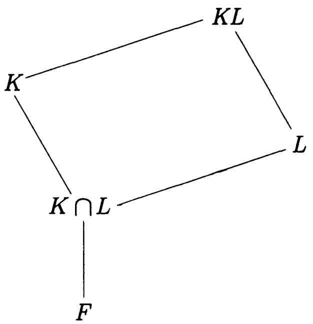
$$\begin{aligned}\sigma^m(\alpha^k) &= (\omega^m\alpha)^k \\ &= \omega^n\alpha^k = \alpha^k.\end{aligned}$$

Moreover, $F(\alpha^k)$ is the fixed field of $\langle\sigma^m\rangle$ for, if m' is a divisor of n and $\sigma^{m'}(\alpha^k) = \alpha^k$, then $\omega^{m'k}\alpha^k = \alpha^k$, which forces n to divide $m'k$. But, $n = mk$, so m divides m', and thus $\sigma^{m'} \in \langle\sigma^m\rangle$. This proves that $\mathrm{Gal}(K/F(\alpha^k)) = \langle\sigma^m\rangle$, so the fundamental theorem tells us that $F(\alpha^k)$ is the fixed field of $\langle\sigma^m\rangle$. We have thus determined the subgroups of $\mathrm{Gal}(K/F)$ and the intermediate fields of K/F to be

$$\begin{gathered}\{\langle\sigma^m\rangle : m \text{ divides } n\},\\ \{F(\alpha^k) : k \text{ divides } n\},\end{gathered}$$

with the correspondence $F(\alpha^k) \leftrightarrow \langle\sigma^m\rangle$ if $km = n$.

Let K/F be Galois, and let L be any extension field of F with K and L inside some common field. Then KL/L is Galois, since if K is the splitting field of a set of separable polynomials over F, then KL is the splitting field of the same set of polynomials over L, and if $f(x) \in F[x]$ is separable over F, then $f(x)$ is separable over L. The following theorem determines the Galois group of KL/L and the degree of this extension.



Theorem 5.5 (Natural Irrationalities) *Let K be a finite Galois extension of F, and let L be an arbitrary extension of F. Then KL/L is Galois and* $\mathrm{Gal}(KL/L) \cong \mathrm{Gal}(K/K \cap L)$. *Moreover,* $[KL : L] = [K : K \cap L]$.

Proof. Define $\theta : \mathrm{Gal}(KL/L) \longrightarrow \mathrm{Gal}(K/F)$ by $\theta(\sigma) = \sigma|_K$. This map is well defined since K is normal over F, and θ is a group homomorphism. The kernel of θ is $\{\sigma \in \mathrm{Gal}(KL/L) : \sigma|_K = \mathrm{id}\}$. However, if $\sigma \in \ker(\theta)$, then $\sigma|_L = \mathrm{id}$ and $\sigma|_K = \mathrm{id}$. Thus, the fixed field of σ contains both K and L, so it contains KL. That means $\sigma = \mathrm{id}$, so θ is injective. Since the image of θ is a subgroup of $\mathrm{Gal}(K/F)$, this image is equal to $\mathrm{Gal}(K/E)$, where E is the fixed field of this image. We show that $E = K \cap L$. If $a \in K \cap L$, then a is fixed by $\sigma|_K$ for each $\sigma \in \mathrm{Gal}(KL/L)$. Therefore, $a \in E$, so $K \cap L \subseteq E$. For the reverse inclusion, let $a \in E$. Then $a \in K$ and $\sigma|_K(a) = a$ for all $\sigma \in \mathrm{Gal}(KL/L)$. Thus, $\sigma(a) = a$ for all such σ, so $a \in L$. This shows $E \subseteq K \cap L$, and so $E = K \cap L$. We have thus proved that

$$\mathrm{Gal}(KL/L) \cong \mathrm{im}(\theta) = \mathrm{Gal}(K/K \cap L).$$

The degree formula follows immediately from this isomorphism. □

A field extension K/F is called *simple* if $K = F(\alpha)$ for some $\alpha \in K$. The next theorem and its corollaries give some conditions for when an extension is simple.

Theorem 5.6 (Primitive Element Theorem) *A finite extension K/F is simple if and only if there are only finitely many fields L with $F \subseteq L \subseteq K$.*

Proof. We prove this with the assumption that $|F| = \infty$. The case for finite fields requires a different proof, which we will handle in Section 6. Suppose that there are only finitely many intermediate fields of K/F. Since $[K : F] < \infty$, we can write $K = F(\alpha_1, \dots, \alpha_n)$ for some $\alpha_i \in K$. We use induction on n; the case $n = 1$ is trivial. If $L = F(\alpha_1, \dots, \alpha_{n-1})$, then since any field between F and L is an intermediate field of K/F, by induction $L = F(\beta)$ for some β. Then $K = F(\alpha_n, \beta)$. For $a \in F$, set $M_a = F(\alpha_n + a\beta)$, an intermediate field of K/F. Since there are only finitely many intermediate fields of K/F but infinitely many elements of F, there are $a, b \in F$ with $a \neq b$ and $M_a = M_b$. Therefore,

$$\beta = \frac{(\alpha_n + b\beta) - (\alpha_n + a\beta)}{b - a} \in M_b.$$

Hence, $\alpha_n = (\alpha_n + b\beta) - b\beta \in M_b$, so $K = F(\alpha_n, \beta) = M_b$. Thus, K is a simple extension of F.

Conversely, suppose that $K = F(\alpha)$ for some $\alpha \in F$. Let M be a field with $F \subseteq M \subseteq K$. Then $K = M(\alpha)$. Let $p(x) = \min(F, \alpha)$ and $q(x) = \min(M, \alpha) \in M[x]$ Then q divides p in $M[x]$. Suppose that $q(x) = a_0 +$

$a_1x + \cdots + x^r$, and set $M_0 = F(a_0, \ldots, a_{r-1}) \subseteq M$. Then $q \in M_0[x]$, so $\min(M_0, \alpha)$ divides q. Thus,

$$\begin{aligned}[K : M] &= \deg(q) \geq \deg(\min(M_0, \alpha)) = [K : M_0] \\ &= [K : M] \cdot [M : M_0].\end{aligned}$$

This implies that $[M : M_0] = 1$, so $M = M_0$. Therefore, M is determined by q. However, there are only finitely many monic divisors of p in $K[x]$, so there are only finitely many such M. □

Corollary 5.7 *If K/F is finite and separable, then $K = F(\alpha)$ for some $\alpha \in K$.*

Proof. If K is finite and separable over F, then $K = F(\alpha_1, \ldots, \alpha_n)$ for some α_i. Let N be the splitting field over F of $\{\min(F, \alpha_i) : 1 \leq i \leq n\}$. Then N/F is Galois by Theorem 4.9 since each $\min(F, \alpha_i)$ is separable over F. Moreover, $K \subseteq N$. By the fundamental theorem, the intermediate fields of N/F are in 1–1 correspondence with the subgroups of the finite group $\mathrm{Gal}(N/F)$. Any finite group has only finitely many subgroups, so N/F has only finitely many intermediate fields. In particular, K/F has only finitely many intermediate fields. Therefore, $K = F(\alpha)$ for some α by the primitive element theorem. □

Corollary 5.8 *If K/F is finite and F has characteristic 0, then $K = F(\alpha)$ for some α.*

Proof. This corollary follows immediately from the preceding corollary since any finite extension of a field of characteristic 0 is separable. □

The normal closure of a field extension

Let K be an algebraic extension of F. The *normal closure* of K/F is the splitting field over F of the set $\{\min(F, a) : a \in K\}$ of minimal polynomials of elements of K. As we will show below, the normal closure N of the extension K/F is a minimal normal extension of F which contains K. This is reasonable since, for each $a \in K$, the polynomial $\min(F, a)$ splits over any normal extension of F containing K. Therefore, the set $\{\min(F, a) : a \in K\}$ is a minimal set of polynomials which must split in any extension of K that is normal over F. We formalize this in the next result, which gives the basic properties of normal closure.

Proposition 5.9 *Let K be an algebraic extension of F, and let N be the normal closure of K/F.*

1. *The field N is a normal extension of F containing K. Moreover, if M is a normal extension of F with $K \subseteq M \subseteq N$, then $M = N$.*

2. *If $K = F(a_1, \ldots, a_n)$, then N is the splitting field of the polynomials $\min(F, a_1), \ldots, \min(F, a_n)$ over F.*

3. *If K/F is a finite extension, then so is N/F.*

4. *If K/F is separable, then N/F is Galois.*

Proof. Since N is a splitting field over F of a set of polynomials, N is normal over F. It is clear that N contains K. Suppose that M is a normal extension of F with $K \subseteq M \subseteq N$. If $a \in K$, then $a \in M$, so by normality $\min(F, a)$ splits over M. However, if X is the set of roots of the polynomials $\{\min(F, a) : a \in K\}$, we have $N = F(X)$. But since these polynomials split over M, all of the roots of these polynomials lie in M. Thus , $X \subseteq M$, and so $N = F(X) \subseteq M$. Therefore, $M = N$.

For part 2, let $L = F(X)$, where $X \subseteq N$ is the set of roots of the polynomials $\{\min(F, a_i) : 1 \leq i \leq n\}$. Then L is a splitting field over F of this set; hence, $K \subseteq L$ and L/F is normal. By part 1, $L = N$.

For the third part, suppose that $[K : F] < \infty$. Then K is a finitely generated extension of F; say that $K = F(a_1, \ldots, a_n)$. Let $p_i(x) = \min(F, a_i)$. By part 2, N is a splitting field of $\{\min(F, a_i) : 1 \leq i \leq n\}$, a finite set of polynomials. Therefore, $[N : F] < \infty$.

Finally, if K/F is separable, then each polynomial $\min(F, a)$ is separable over F. Therefore, N is the splitting field of the set $\{\min(F, a) : a \in K\}$ of separable polynomials over F, so N is Galois over F. □

The normal closure of an algebraic extension K/F is uniquely determined by the conditions in the first part of the previous proposition, as we now show.

Corollary 5.10 *Let K be an algebraic extension of F, and let N be the normal closure of K/F. If N' is any normal extension of F containing K, then there is an F-homomorphism from N to N'. Consequently, if N' does not contain any proper subfield normal over F that contains K, then N and N' are F-isomorphic.*

Proof. Suppose that N' is normal over F and contains K. Then $\min(F, a)$ splits over N' for each $a \in K$. By the isomorphism extension theorem, the identity map on F extends to a homomorphism $\sigma : N \to N'$. Then $\sigma(N)$ is a splitting field of $\{\min(F, a) : a \in K\}$ in N', so $\sigma(N)$ is normal over F and contains K. Therefore, if N' does not contain any proper subfield normal over F that contains K, then $\sigma(N) = N'$, so N and N' are F-isomorphic. □

Example 5.11 Let $F = \mathbb{Q}$ and $K = \mathbb{Q}(\sqrt[3]{2})$. If $\omega^3 = 1$ and $\omega \neq 1$, then $\mathbb{Q}(\sqrt[3]{2}, \omega)$ is the splitting field of $x^3 - 2$ over $\mathbb{Q}$, so it is normal over $\mathbb{Q}$. This field is clearly the smallest extension of K that is normal over $\mathbb{Q}$, so $\mathbb{Q}(\sqrt[3]{2}, \omega)$ is the normal closure of $\mathbb{Q}(\omega)/\mathbb{Q}$.

Example 5.12 If K is an extension of F, and if $a \in K$ has minimal polynomial $p(x)$ over F, then the normal closure of $F(a)/F$ is the field $F(a_1, a_2, \ldots, a_n)$, where the a_i are the roots of $p(x)$.

Suppose that K/F is a finite separable extension with normal closure N. Let $G = \text{Gal}(N/F)$ and $H = \text{Gal}(N/K)$. So $K = \mathcal{F}(H)$. Suppose that K is not Galois over F. Then H is not normal in G. The minimality of N as a normal extension of F containing K translates via the fundamental theorem into the following group theoretic relation between G and H: The largest normal subgroup of G contained in H is $\langle \text{id} \rangle$ for, if $H' \subseteq H$ is a normal subgroup of G, then $L = \mathcal{F}(H')$ is an extension of K that is normal over F. But, as $L \subseteq N$, minimality of N implies that $L = N$, so $H' = \langle \text{id} \rangle$. Recall from group theory that if H is a subgroup of a group G, then $\bigcap_{g \in G} gHg^{-1}$ is the largest normal subgroup of a group G contained in a subgroup H. Therefore, in the context above, $\bigcap_{g \in G} gHg^{-1} = \langle \text{id} \rangle$.

The fundamental theorem of algebra

The fundamental theorem of algebra states that every polynomial in $\mathbb{C}[x]$ has a root in $\mathbb{C}$. This was first proved by Gauss and is commonly proved using the theory of analytic functions in a course in complex analysis. We give here a proof using Galois theory, which combines the fundamental theorem and the Sylow theorems of group theory. It is a nice application of the interaction of group and field theory.

To prove the fundamental theorem of algebra, we do need to know one result from analysis, namely the intermediate value theorem. Beyond this, we can give a proof using group theory and Galois theory. We point out the group theoretic fact we need: If G is a finite group whose order is a power of a prime p, then any maximal subgroup of G has index p in G. This fact can be found in Proposition 2.4 of Appendix C.

Lemma 5.13 *Let* $f(x) \in \mathbb{R}[x]$.

1. *If* $f(x) = x^2 - a$ *for some* $a > 0$, *then* f *has a root in* $\mathbb{R}$. *Therefore, every nonnegative real number has a real square root.*

2. *If* $\deg(f)$ *is odd, then* f *has a root in* $\mathbb{R}$. *Consequently, the only odd degree extension of* $\mathbb{R}$ *is* $\mathbb{R}$ *itself.*

Proof. Suppose that $f(x) = x^2 - a$ with $a > 0$. Then $f(0) < 0$ and $f(u) > 0$ for u sufficiently large. Therefore, there is a $c \in [0, u]$ with $f(c) = 0$ by the intermediate value theorem. In other words, $\sqrt{a} = c \in \mathbb{R}$.

For part 2, suppose that the leading coefficient of f is positive. Then

$$\lim_{x \to \infty} f(x) = \infty \quad \text{and} \quad \lim_{x \to -\infty} f(x) = -\infty.$$

By another use of the intermediate value theorem, there is a $c \in \mathbb{R}$ with $f(c) = 0$. If $L/\mathbb{R}$ is an odd degree extension, take $a \in L - \mathbb{R}$. Then $\mathbb{R}(a)/\mathbb{R}$ is

also of odd degree, so $\deg(\min(\mathbb{R}, a))$ is odd. However, this polynomial has a root in $\mathbb{R}$ by what we have just shown. Since this polynomial is irreducible, this forces $\min(\mathbb{R}, a)$ to be linear, so $a \in \mathbb{R}$. Therefore, $L = \mathbb{R}$.

Lemma 5.14 *Every complex number has a complex square root. Therefore, there is no field extension N of $\mathbb{C}$ with $[N : \mathbb{C}] = 2$.*

Proof. To prove this, we use the polar coordinate representation of complex numbers. Let $a \in \mathbb{C}$, and set $a = re^{i\theta}$ with $r \geq 0$. Then $\sqrt{r} \in \mathbb{R}$ by Lemma 5.13, so $b = \sqrt{r}e^{i\theta/2} \in \mathbb{C}$. We have $b^2 = r(e^{i\theta/2})^2 = re^{i\theta} = a$. If N is an extension of $\mathbb{C}$ with $[N : \mathbb{C}] = 2$, then there is an $a \in \mathbb{C}$ with $N = \mathbb{C}(\sqrt{a})$. But, the first part of the lemma shows that $\mathbb{C}(\sqrt{a}) = \mathbb{C}$, so there are no quadratic extensions of $\mathbb{C}$. □

Theorem 5.15 (Fundamental Theorem of Algebra) *The field $\mathbb{C}$ is algebraically closed.*

Proof. Let L be a finite extension of $\mathbb{C}$. Since $\text{char}(\mathbb{R}) = 0$, the field L is separable over $\mathbb{R}$, and L is also a finite extension of $\mathbb{R}$. Let N be the normal closure of $L/\mathbb{R}$. We will show that $N = \mathbb{C}$, which will prove the theorem. Let $G = \text{Gal}(N/\mathbb{R})$. Then

$$\begin{aligned}|G| &= [N : \mathbb{R}] = [N : \mathbb{C}] \cdot [\mathbb{C} : \mathbb{R}] \\ &= 2[N : \mathbb{C}]\end{aligned}$$

is even. Let H be a 2-Sylow subgroup of G, and let E be the fixed field of H. Then $|G : H| = [E : \mathbb{R}]$ is odd. Thus, by Lemma 5.13, we see that $E = F$, so $G = H$ is a 2-group. Therefore, $\text{Gal}(N/\mathbb{C})$ is also a 2-group. Let P be a maximal subgroup of $\text{Gal}(N/\mathbb{C})$. By the theory of p-groups, $[\text{Gal}(N/\mathbb{C}) : P] = 2$. If T is the fixed field of M, then $[T : \mathbb{C}] = 2$. This is impossible by Lemma 5.14. This contradiction shows that $|G| = 1$, so $N = \mathbb{C}$. □

Problems

1. A *transitive* subgroup of S_n is a subgroup G such that for each $i, j \in \{1, \ldots, n\}$, there is a $\sigma \in G$ with $\sigma(i) = j$. If K is the splitting field over F of a separable irreducible polynomial $f(x) \in F[x]$ of degree n, show that $|\text{Gal}(K/F)|$ is divisible by n and that $\text{Gal}(K/F)$ is isomorphic to a transitive subgroup of S_n. Conclude that $[K : F]$ divides $n!$.

2. Write down all the transitive subgroups of S_3 and S_4.

3. Determine all the transitive subgroups G of S_5 for which $|G|$ is a multiple of 5. For each transitive subgroup, find a field F and an irreducible polynomial of degree 5 over F such that if K is the splitting

field of f over F, then $\mathrm{Gal}(K/F)$ is isomorphic to the given subgroup. (Hint: This will require use of semidirect products.)

4. In the following problems, let K be the splitting field of $f(x)$ over F. Determine $\mathrm{Gal}(K/F)$ and find all the intermediate subfields of K/F.

   (a) $F = \mathbb{Q}$ and $f(x) = x^4 - 7$.
   (b) $F = \mathbb{F}_5$ and $f(x) = x^4 - 7$.
   (c) $F = \mathbb{Q}$ and $f(x) = x^5 - 2$.
   (d) $F = \mathbb{F}_2$ and $f(x) = x^6 + 1$.
   (e) $F = \mathbb{Q}$ and $f(x) = x^8 - 1$.

5. Let K be a Galois extension of F with $[K : F] = n$. If p is a prime divisor of n, show that there is a subfield L of K with $[K : L] = p$.

6. Let N be a Galois extension of F with $\mathrm{Gal}(N/F) = A_4$. Show that there is no intermediate field of N/F with $[N : F] = 2$.

7. Give examples of field extensions K/F with

   (a) K/F normal but not Galois,
   (b) K/F separable but not Galois.

8. Let K/F be Galois with $G = \mathrm{Gal}(K/F)$, and let L be an intermediate field. Let $N \subseteq K$ be the normal closure of L/F. If $H = \mathrm{Gal}(K/L)$, show that $\mathrm{Gal}(K/N) = \bigcap_{\sigma \in G} \sigma H \sigma^{-1}$.

9. Let K be a Galois extension of F and let $a \in K$. Let $n = [K : F]$, $r = [F(a) : F]$, and $H = \mathrm{Gal}(K/F(a))$. Let $\tau_1, \ldots, \tau_r$ be left coset representatives of H in G. Show that $\min(F, a) = \prod_{i=1}^{r}(x - \tau_i(a))$. Conclude that
$$\prod_{\sigma \in \mathrm{Gal}(K/F)} (x - \sigma(a)) = \min(F, a)^{n/r}.$$

10. Let K be a Galois extension of F, and let $a \in K$. Let $L_a : K \to K$ be the F-linear transformation defined by $L_a(b) = ab$. Show that the characteristic polynomial of L_a is equal to $\prod_{\sigma \in \mathrm{Gal}(K/F)}(x - \sigma(a))$ and the minimal polynomial of L_a is $\min(F, a)$.

11. Let K be a finite Galois extension of F with Galois group G. Let L be an intermediate extension, and let H be the corresponding subgroup of G. If $N(H)$ is the normalizer of H in G, let L_0 be the fixed field of $N(H)$. Show that L/L_0 is Galois and that if M is any subfield of L containing F for which L/M is Galois, then M contains L_0.

12. Let F be a field of characteristic not 2, and let K be a Galois extension with $[K : F] = 4$. Prove that if $\text{Gal}(K/F) \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, then $K = F(\sqrt{a}, \sqrt{b})$ for some $a, b \in F$.

13. If K is the splitting field of $x^4 - 2$ over $\mathbb{Q}$, find $\text{Gal}(K/\mathbb{Q})$ and find all intermediate subfields. To what group is $\text{Gal}(K/\mathbb{Q})$ abstractly isomorphic?

14. If K is the splitting field of $x^5 - 11$ over $\mathbb{Q}$, find $\text{Gal}(K/\mathbb{Q})$ and find all intermediate subfields.

15. Let K be a finite normal extension of F such that there are no proper intermediate extensions of K/F. Show that $[K : F]$ is prime. Give a counterexample if K is not normal over F.

16. Let K be a Galois extension of $\mathbb{Q}$. View K as a subfield of $\mathbb{C}$. If σ is complex conjugation, show that $\sigma(K) = K$, so $\sigma|_K \in \text{Gal}(K/\mathbb{Q})$. Show that $\mathcal{F}(\sigma|_K) = K \cap \mathbb{R}$, and conclude that $[K : K \cap \mathbb{R}] \leq 2$. Give examples to show that both $[K : K \cap \mathbb{R}] = 1$ and $[K : K \cap \mathbb{R}] = 2$ can occur.

17. Prove the *normal basis theorem*: If K is a finite Galois extension of F, then there is an $a \in K$ such that $\{\sigma(a) : \sigma \in \text{Gal}(K/F)\}$ is a basis for K as an F-vector space.

18. Let Q_8 be the *quaternion group* $\{\pm 1, \pm i, \pm j, \pm k\}$, where multiplication is determined by the relations $i^2 = j^2 = -1$ and $ij = k = -ji$. Show that Q_8 is not isomorphic to a subgroup of S_4. Conclude that Q_8 is not the Galois group of the splitting field of a degree 4 polynomial over a field.

19. (a) Let $K \subseteq N$ both be Galois extensions of a field F. Show that the map $\varphi : \text{Gal}(N/F) \to \text{Gal}(K/F)$ given by $\varphi(\sigma) = \sigma|_K$ is a surjective group homomorphism. Therefore, $\text{Gal}(K/F) = \{\sigma|_K : \sigma \in \text{Gal}(N/F)\}$. Show that $\ker(\varphi) = \text{Gal}(N/K)$.
    (b) Let K and L be Galois extensions of F. Show that the restriction of function map defined in (a) induces an injective group homomorphism $\text{Gal}(KL/F) \to \text{Gal}(K/F) \oplus \text{Gal}(L/F)$. Show that this map is surjective if and only if $K \cap L = F$.

20. Let k be a field of characteristic $p > 0$, let $K = k(x, y)$ be the rational function field in two variables over k, and let $F = k(x^p, y^p)$.
    (a) Prove that $[K : F] = p^2$.
    (b) Prove that $K^p \subseteq F$.
    (c) Prove that there is no $\alpha \in K$ with $K = F(\alpha)$.
    (d) Exhibit an infinite number of intermediate fields of K/F.

21. This problem gives an alternative proof of the primitive element theorem for infinite fields.

   (a) Let V be a finite dimensional F-vector space, where F is an infinite field. Show that V is not the union of finitely many proper subspaces.

   (b) Let K/F be a finite extension of finite fields. Show that K is not the union of the proper intermediate fields of K/F. Conclude that if $\{K_i\}$ is the set of proper intermediate fields and $a \in K - \bigcup K_i$, then $K = F(a)$.

22. Let $K = \mathbb{Q}(X)$, where $X = \{\sqrt{p} : p \text{ is prime}\}$. Show that K is Galois over $\mathbb{Q}$. If $\sigma \in \text{Gal}(K/\mathbb{Q})$, let $Y_\sigma = \{\sqrt{p} : \sigma(\sqrt{p}) = -\sqrt{p}\}$. Prove the following statements.

   (a) If $Y_\sigma = Y_\tau$, then $\sigma = \tau$.

   (b) If $Y \subseteq X$, then there is a $\sigma \in \text{Gal}(K/\mathbb{Q})$ with $Y_\sigma = Y$.

   (c) If $\mathcal{P}(X)$ is the power set of X, show that $|\text{Gal}(K/\mathbb{Q})| = |\mathcal{P}(X)|$ and that $|X| = [K : \mathbb{Q}]$, and conclude that $|\text{Gal}(K/\mathbb{Q})| > [K : \mathbb{Q}]$.

   (Hint: A Zorn's lemma argument may help in (b). You may want to verify that if $Y \subseteq X$ and $\sqrt{p} \notin Y$, then $[\mathbb{Q}(Y)(\sqrt{p}) : \mathbb{Q}(Y)] = 2$. The inequality $|\mathcal{P}(X)| > |X|$ is proved in Example 2.2 of Appendix B.)

23. Suppose that K is an extension of F with $[K : F] = 2$. If $\text{char}(F) \neq 2$, show that K/F is Galois.

24. Let $F \subseteq L \subseteq K$ be fields such that L/F is purely inseparable. Let $a \in K$ be separable over F. Prove that $\min(F, a) = \min(L, a)$. Use this to prove the following statement: Suppose that $F \subseteq L \subseteq K$ are fields such that L/F is purely inseparable, K/L is separable, and $[K : F] < \infty$. Let S be the separable closure of F in K. Then $K = SL$ and $[K : L] = [S : F]$.

25. This problem outlines a proof that the separable degree $[K : F]_s$ of a finite extension K/F is equal to the number of F-homomorphisms from K to an algebraic closure of F.

   (a) Suppose that $K = F(a)$, and let $f(x) = \min(F, a)$. If N is an algebraic closure of F and $b \in N$ is a root of f, show that there is an F-homomorphism $K \to N$ that sends a to b.

   (b) If $K = F(a)$ as above, show that all F-homomorphisms from K to N are obtained in the manner of the previous step. Conclude that $[K : F]_s$ is equal to the number of such F-homomorphisms.

(c) Let K/F be a finite extension, and let S be the separable closure of F in K. Show that any F-homomorphism from S to N extends uniquely to K. Use the previous step to conclude that $[S : F] = [K : F]_s$ is the number of F-homomorphisms from K to N.

26. Let K/F be a normal extension and let L/F be an algebraic extension. If either K/F or L/F is separable, show that $[KL : L] = [K : K \cap L]$. Give an example to show that this can be false without the separability hypothesis.

27. Let F be a field. Show that the rational function field $F(x)$ is not algebraically closed.

28. Let F be a finite extension of $\mathbb{Q}$. Show that F is not algebraically closed.

II
Some Galois Extensions

Now that we have developed the machinery of Galois theory, we apply it in this chapter to study special classes of field extensions. Sections 9 and 11 are good examples of how we can use group theoretic information to obtain results in field theory. Section 10 has a somewhat different flavor than the other sections. In it, we look into the classical proof of the Hilbert Theorem 90, a result originally used to help describe cyclic extensions, and from that proof we are led to the study of cohomology, a key tool in algebraic topology, algebraic geometry, and the theory of division rings.

6 Finite Fields

In this section, we study finite fields and, more generally, finite extensions of finite fields.

Let F be a finite field, and say $\text{char}(F) = p$. We can view F as an extension field of $\mathbb{F}_p$. Since F is finite, F is a finite dimensional $\mathbb{F}_p$-vector space. If $[F : \mathbb{F}_p] = n$, then F and $\mathbb{F}_p$ are isomorphic as $\mathbb{F}_p$-vector spaces, so $|F| = p^n$. We will first obtain some field theoretic information about F by investigating the group structure of the multiplicative group F^*. For the next lemma, recall that if G is an Abelian group, then the *exponent* $\exp(G)$ of G is the least common multiple of elements in G. By a group theory exercise, there is an element of G whose order is $\exp(G)$. From this fact, it follows that G is cyclic if and only if $|G| = \exp(G)$. These facts are proven in Proposition 1.4 of Appendix C.

Lemma 6.1 *If K is a field and G is a finite subgroup of K^*, then G is cyclic.*

Proof. Let $n = |G|$ and $m = \exp(G)$. Then m divides n by Lagrange's theorem. If $g \in G$, then $g^m = 1$, so each element of G is a root of the polynomial $x^m - 1$. This polynomial has at most m roots in the field K. However, $x^m - 1$ has at least the elements of G as roots, so $n \leq m$. Therefore, $\exp(G) = |G|$, so G is cyclic. □

Corollary 6.2 *If F is a finite field, then F^* is cyclic.*

Example 6.3 Let $F = \mathbb{F}_p$. A generator for F^* is often called a *primitive root modulo p*. For example, 2 is a primitive root modulo 5. Moreover, 2 is not a primitive root modulo 7, while 3 is a primitive root modulo 7. In general, it is not easy to find a primitive root modulo p, and there is no simple way to find a primitive root in terms of p.

In Section 5, the primitive element theorem was stated for arbitrary base fields but was proved only for infinite fields. If K/F is an extension of finite fields, then there are finitely many intermediate fields. Therefore, the hypotheses of the primitive element theorem hold for K/F. The following corollary finishes the proof of the primitive element theorem.

Corollary 6.4 *If K/F is an extension of finite fields, then K is a simple extension of F.*

Proof. By the previous corollary, the group K^* is cyclic. Let α be a generator of the cyclic group K^*. Every nonzero element of K is a power of α, so $K = F(\alpha)$. Therefore, K is a simple extension of F. □

The following theorem exploits group theoretic properties of finite groups to give the main structure theorem of finite fields.

Theorem 6.5 *Let F be a finite field with $\operatorname{char}(F) = p$, and set $|F| = p^n$. Then F is the splitting field of the separable polynomial $x^{p^n} - x$ over $\mathbb{F}_p$. Thus, $F/\mathbb{F}_p$ is Galois. Furthermore, if σ is defined on F by $\sigma(a) = a^p$, then σ generates the Galois group $\operatorname{Gal}(F/\mathbb{F}_p)$, so this Galois group is cyclic.*

Proof. Let $|F| = p^n$, so $|F^*| = p^n - 1$. By Lagrange's theorem, if $a \in F^*$, then $a^{p^n-1} = 1$. Multiplying by a gives $a^{p^n} = a$. This equation also holds for $a = 0$. Therefore, the elements of F are roots of the polynomial $x^{p^n} - x$. However, this polynomial has at most p^n roots, so the elements of F are precisely the roots of $x^{p^n} - x$. This proves that F is the splitting field over $\mathbb{F}_p$ of $x^{p^n} - x$, and so F is normal over $\mathbb{F}_p$. Moreover, the derivative test shows that $x^{p^n} - x$ has no repeated roots, so $x^{p^n} - x$ is separable over $\mathbb{F}_p$. Thus, F is Galois over $\mathbb{F}_p$.

Define $\sigma : F \longrightarrow F$ by $\sigma(a) = a^p$. An easy computation shows that σ is an $\mathbb{F}_p$-homomorphism, and σ is surjective since F is finite. Hence, σ is an $\mathbb{F}_p$-automorphism of F. The fixed field of σ is $\{a \in F : a^p = a\} \supseteq \mathbb{F}_p$. Each element in $\mathcal{F}(\sigma)$ is a root of $x^p - x$, so there are at most p elements in $\mathcal{F}(\sigma)$. This proves that $\mathbb{F}_p = \mathcal{F}(\sigma)$, so $\mathrm{Gal}(F/\mathbb{F}_p)$ is the cyclic group generated by σ. □

The automorphism σ defined above is called the *Frobenius automorphism* of F.

Corollary 6.6 *Any two finite fields of the same size are isomorphic.*

Proof. The proof of Theorem 6.5 shows that any two fields of order p^n are splitting fields over $\mathbb{F}_p$ of $x^{p^n} - x$, so the corollary follows from the isomorphic extension theorem. □

We can use Theorem 6.5 to describe any finite extension of finite fields, not only extensions of $\mathbb{F}_p$.

Corollary 6.7 *If K/F is an extension of finite fields, then K/F is Galois with a cyclic Galois group. Moreover, if* $\mathrm{char}(F) = p$ *and* $|F| = p^n$*, then* $\mathrm{Gal}(K/F)$ *is generated by the automorphism τ defined by $\tau(a) = a^{p^n}$.*

Proof. Say $[K : \mathbb{F}_p] = m$. Then $\mathrm{Gal}(K/\mathbb{F}_p)$ is a cyclic group of order m by Theorem 6.5, so the order of the Frobenius automorphism σ of K is m. The group $\mathrm{Gal}(K/F)$ is a subgroup of $\mathrm{Gal}(K/\mathbb{F}_p)$, so it is also cyclic. If $s = |\mathrm{Gal}(K/F)|$ and $m = ns$, then a generator of $\mathrm{Gal}(K/F)$ is σ^n. By induction, we see that the function σ^n is given by $\sigma^n(a) = a^{p^n}$. Also, since $s = [K : F]$, we have that $n = [F : \mathbb{F}_p]$, so $|F| = p^n$. □

We have described finite fields as extensions of $\mathbb{F}_p$ and have shown that any finite extension of $\mathbb{F}_p$ has p^n elements for some n. However, we have not yet determined for which n there is a field with p^n elements. Using the fundamental theorem along with the description of finite fields as splitting fields in Theorem 6.5, we now show that for each n there is a unique up to isomorphism field with p^n elements.

Theorem 6.8 *Let N be an algebraic closure of $\mathbb{F}_p$. For any positive integer n, there is a unique subfield of N of order p^n. If K and L are subfields of N of orders p^m and p^n, respectively, then $K \subseteq L$ if and only if m divides n. When this occurs, L is Galois over K with Galois group generated by τ, where $\tau(a) = a^{p^n}$.*

Proof. Let n be a positive integer. The set of roots in N of the polynomial $x^{p^n} - x$ has p^n elements and is a field. Thus, there is a subfield of N of order p^n. Since any two fields of order p^n in N are splitting fields of $x^{p^n} - x$

over $\mathbb{F}_p$ by Theorem 6.5, any subfield of N of order p^n consists exactly of the roots of $x^{p^n} - x$. Therefore, there is a unique subfield of N of order p^n.

Let K and L be subfields of N, of orders p^m and p^n, respectively. First, suppose that $K \subseteq L$. Then

$$\begin{aligned} n = [L : \mathbb{F}_p] &= [L : K] \cdot [K : \mathbb{F}_p] \\ &= m[L : K], \end{aligned}$$

so m divides n. Conversely, suppose that m divides n. Each element a of K satisfies $a^{p^m} = a$. Since m divides n, each a also satisfies $a^{p^n} = a$, so $a \in L$. This proves that $K \subseteq L$. When this happens L is Galois over K by Corollary 6.7. That corollary also shows that $\mathrm{Gal}(L/K)$ is generated by τ, where τ is defined by $\tau(a) = a^{|K|}$. □

If F is a finite field and $f(x) \in F[x]$, then Theorems 6.5 and 6.8 can be used to determine the splitting field over F of the polynomial f.

Corollary 6.9 *Let F be a finite field, and let $f(x)$ be a monic irreducible polynomial over F of degree n.*

1. *If a is a root of f in some extension field of F, then $F(a)$ is a splitting field for f over F. Consequently, if K is a splitting field for f over F, then $[K : F] = n$.*

2. *If $|F| = q$, then the set of roots of f is $\{a^{q^r} : r \geq 1\}$.*

Proof. Let K be a splitting field of f over F. If $a \in K$ is a root of $f(x)$, then $F(a)$ is an n-dimensional extension of F inside K. By Theorem 6.5, $F(a)$ is a Galois extension of F; hence, $f(x) = \min(F, a)$ splits over $F(a)$. Therefore, $F(a)$ is a splitting field of f over F, so $K = F(a)$. This proves the first statement. For the second, we note that $\mathrm{Gal}(K/F) = \langle \sigma \rangle$, where $\sigma(c) = c^q$ for any $c \in K$, by Theorem 6.8. Each root of f is then of the form $\sigma^r(a) = a^{q^r}$ by the isomorphism extension theorem, which shows that the set of roots of f is $\{a^{q^r} : r \geq 1\}$. □

Example 6.10 Let $F = \mathbb{F}_2$ and $K = F(\alpha)$, where α is a root of $f(x) = x^3 + x^2 + 1$. This polynomial has no roots in F, as a quick calculation shows, so it is irreducible over F and $[K : F] = 3$. The field K is the splitting field of f over F, and the roots of f are α, α^2, and α^4, by Corollary 6.9. Since $f(\alpha) = 0$, we see that $\alpha^3 = \alpha^2 + 1$, so $\alpha^4 = \alpha^3 + \alpha = \alpha^2 + \alpha + 1$. Therefore, in terms of the basis $\{1, \alpha, \alpha^2\}$ for K/F, the roots of f are α, α^2, and $1 + \alpha + \alpha^2$. This shows explicitly that $F(\alpha)$ is the splitting field of f over F.

Example 6.11 Let $F = \mathbb{F}_2$ and $f(x) = x^4 + x + 1$. By the derivative test, we see that f has no repeated roots. The polynomial f is irreducible over

f, since f has no roots in F and is not divisible by the unique irreducible quadratic $x^2 + x + 1$ in $F[x]$. If α is a root of f, then $\alpha^4 = \alpha + 1$; hence, the roots of f are α, $\alpha + 1$, α^2, and $\alpha^2 + 1$.

Example 6.12 Let $f(x) = x^2 + 1$. If p is an odd prime, then we show that f is reducible over $F = \mathbb{F}_p$ if and only if $p \equiv 1 (\bmod 4)$. To prove this, if $a \in F$ is a root of $x^2 + 1$, then $a^2 = -1$, so a has order 4 in F^*. By Lagrange's theorem, 4 divides $|F^*| = p - 1$, so $p \equiv 1 (\bmod 4)$. Conversely, if $p \equiv 1 (\bmod 4)$, then 4 divides $p - 1$, so there is an element $a \in F^*$ of order 4, since F^* is a cyclic group of order $p - 1$. Thus, $a^4 = 1$ and $a^2 \neq 1$. This forces $a^2 = -1$, so a is a root of f.

If F is a finite field, then we have seen that every finite extension of F is Galois over F. Hence, every extension of F is separable over F. Every algebraic extension of F is then separable over F, so F is perfect. To note this more prominently, we record this as a corollary. We have already seen this fact in Example 4.14.

Corollary 6.13 *Every finite field is perfect.*

Given an integer n, Theorem 6.8 shows that there is a finite field with p^n elements. For a specific n, how do we go about finding this field? To construct finite fields, we can use irreducible polynomials over $\mathbb{F}_p$. Note that if $f(x)$ is an irreducible polynomial of degree n in $\mathbb{F}_p[x]$, then $\mathbb{F}_p[x]/(f(x))$ is a field extension of degree n over $\mathbb{F}_p$; hence, it has p^n elements. Conversely, if F has p^n elements, and if $F = \mathbb{F}_p(\alpha)$, then $\min(\mathbb{F}_p, \alpha)$ is an irreducible polynomial of degree n. Therefore, finding finite fields is equivalent to finding irreducible polynomials in $\mathbb{F}_p[x]$. For instance, $\mathbb{Z}_2[x]/(x^2 + x + 1)$ is a field of 4 elements, and $\mathbb{Z}_5[x]/(x^4 - 7)$ is a field of $5^4 = 625$ elements. The following proposition gives one way of searching for irreducible polynomials over $\mathbb{F}_p$.

Proposition 6.14 *Let n be a positive integer. Then $x^{p^n} - x$ factors over $\mathbb{F}_p$ into the product of all monic irreducible polynomials over $\mathbb{F}_p$ of degree a divisor of n.*

Proof. Let F be a field of order p^n. Then F is the splitting field of $x^{p^n} - x$ over $\mathbb{F}_p$ by Theorem 6.5. Recall that F is exactly the set of roots of $x^{p^n} - x$. Let $a \in F$, and set $m = [\mathbb{F}_p(a) : \mathbb{F}_p]$, a divisor of $[F : \mathbb{F}_p]$. The polynomial $\min(\mathbb{F}_p, a)$ divides $x^{p^n} - x$, since a is a root of $x^{p^n} - x$. Conversely, if $f(x)$ is a monic irreducible polynomial over $\mathbb{F}_p$ of degree m, where m divides n, let K be the splitting field of f over $\mathbb{F}_p$ inside some algebraic closure of F. If a is a root of f in K, then $K = \mathbb{F}_p(a)$ by Corollary 6.9. Therefore, $[K : \mathbb{F}_p] = m$, so $K \subseteq F$ by Theorem 6.8. Thus, $a \in F$, so a is a root of $x^{p^n} - x$. Since f is irreducible over $\mathbb{F}_p$, we have $f = \min(\mathbb{F}_p, a)$, so f divides $x^{p^n} - x$. Since $x^{p^n} - x$ has no repeated roots, $x^{p^n} - x$ factors into distinct

irreducible factors over $\mathbb{F}_p$. We have shown that the irreducible factors of $x^{p^n} - x$ are exactly the irreducible polynomials of degree a divisor of n; hence, the proposition is proven. □

Example 6.15 The monic irreducible polynomials of degree 5 over $\mathbb{F}_2$ can be determined by factoring $x^{2^5} - x$, which we see factors as

$$\begin{aligned} x^{2^5} - x = x\,(x+1)\left(x^5+x^3+1\right)\left(x^5+x^2+1\right) \\ \times\left(x^5+x^4+x^3+x+1\right)\left(x^5+x^4+x^2+x+1\right) \\ \times\left(x^5+x^4+x^3+x^2+1\right)\left(x^5+x^3+x^2+x+1\right). \end{aligned}$$

This factorization produces the six monic irreducible polynomials of degree 5 over $\mathbb{F}_2$. Note that we only need one of these polynomials in order to construct a field with 2^5 elements. Similarly, the monic irreducible polynomials of degree 2, 3, or 6 over $\mathbb{F}_2$ can be found by factoring $x^{2^6} - x$. For example, $x^6 + x + 1$ is an irreducible factor of $x^{64} - x$, so $\mathbb{F}_2[x]/(x^6 + x + 1)$ is a field with 64 elements. The factorization of $x^{32} - x$ and the factor $x^6 + x + 1$ of $x^{64} - x$ was found by using the computer algebra program Scientific Workplace.

Problems

1. Let G be a finite Abelian group.

   (a) If $a, b \in G$ have orders n and m, respectively, and if $\gcd(n, m) = 1$, show that ab has order nm.

   (b) If $a, b \in G$ have orders n and m, respectively, show that there is an element of G whose order is $\mathrm{lcm}(a, b)$.

   (c) Show that there is an element of G whose order is $\exp(G)$.

2. Let p be a prime, and let F be a field with $|F| = p^2$. Show that there is an $a \in F$ with $a^2 = 5$. Generalize this statement, and prove the generalization.

3. Let F be a finite field. Prove that there is an irreducible polynomial of degree n over F for any n.

4. Let K be a field with $|K| = 4$. Show that $K = \mathbb{F}_2(\alpha)$, where $\alpha^2 + \alpha + 1 = 0$.

5. Determine the irreducible factorization of $x^4 + 1$ over $\mathbb{F}_3$.

6. Let F be a finite field. If $f, g \in F[x]$ are irreducible polynomials of the same degree, show that they have the same splitting field. Use this to determine the splitting field of $x^4 + 1$ over $\mathbb{F}_3$.

7. Let q be a power of a prime p, and let n be a positive integer not divisible by p. We let $\mathbb{F}_q$ be the unique up to isomorphism finite field of q elements. If K is the splitting field of $x^n - 1$ over $\mathbb{F}_q$, show that $K = \mathbb{F}_{q^m}$, where m is the order of q in the group of units $(\mathbb{Z}/n\mathbb{Z})^*$ of the ring $\mathbb{Z}/n\mathbb{Z}$.

8. Let F be a field of characteristic p.

   (a) Let $F^p = \{a^p : a \in F\}$. Show that F^p is a subfield of F.

   (b) If $F = \mathbb{F}_p(x)$ is the rational function field in one variable over $\mathbb{F}_p$, determine F^p and $[F : F^p]$.

9. Show that $x^4 - 7$ is irreducible over $\mathbb{F}_5$.

10. Show that every element of a finite field is a sum of two squares.

11. Let F be a field with $|F| = q$. Determine, with proof, the number of monic irreducible polynomials of prime degree p over F, where p need not be the characteristic of F.

12. Let K and L be extensions of a finite field F of degrees n and m, respectively. Show that KL has degree $\operatorname{lcm}(n, m)$ over F and that $K \cap L$ has degree $\gcd(n, m)$ over F.

13. (a) Show that $x^3 + x^2 + 1$ and $x^3 + x + 1$ are irreducible over $\mathbb{F}_2$.

    (b) Give an explicit isomorphism between $\mathbb{F}_2[x]/(x^3 + x^2 + 1)$ and $\mathbb{F}_2[x]/(x^3 + x + 1)$.

14. Let k be the algebraic closure of $\mathbb{Z}_p$, and let $\varphi \in \operatorname{Gal}(k/\mathbb{Z}_p)$ be the Frobenius map $\varphi(a) = a^p$. Show that φ has infinite order, and find a $\sigma \in \operatorname{Gal}(k/\mathbb{Z}_p)$ with $\sigma \notin \langle\varphi\rangle$.

15. Let N be an algebraic closure of a finite field F. Prove that $\operatorname{Gal}(N/F)$ is an Abelian group and that any automorphism in $\operatorname{Gal}(N/F)$ is of infinite order.
    (By techniques of infinite Galois theory, one can prove that $\operatorname{Gal}(N/\mathbb{F}_p)$ is isomorphic to the additive group of the p-adic integers; see Section 17.)

7 Cyclotomic Extensions

An nth root of unity is an element ω of a field with $\omega^n = 1$. For instance, the complex number $e^{2\pi i/n}$ is an nth root of unity. We have seen roots of unity arise in various examples. In this section, we investigate the field extension $F(\omega)/F$, where ω is an nth root of unity. Besides being interesting extensions in their own right, these extensions will play a role in

applications of Galois theory to ruler and compass constructions and to the question of solvability of polynomial equations.

Definition 7.1 *If* $\omega \in F$ *with* $\omega^n = 1$*, then* ω *is an nth root of unity. If the order of* ω *is* n *in the multiplicative group* F^**, then* ω *is a primitive nth root of unity. If* ω *is any root of unity, then the field extension* $F(\omega)/F$ *is called a cyclotomic extension.*

We point out two facts about roots of unity. First, if $\omega \in F$ is a primitive nth root of unity, then we see that $\operatorname{char}(F)$ does not divide n for, if $n = pm$ with $\operatorname{char}(F) = p$, then $0 = \omega^n - 1 = (\omega^m - 1)^p$. Therefore, $\omega^m = 1$, and so the order of ω is not n. Second, if ω is an nth root of unity, then the order of ω in the group F^* divides n, so the order of ω is equal to some divisor m of n. The element ω is then a primitive mth root of unity.

The nth roots of unity in a field K are exactly the set of roots of $x^n - 1$. Suppose that $x^n - 1$ splits over K, and let G be the set of roots of unity in K. Then G is a finite subgroup of K^*, so G is cyclic by Lemma 6.1. Any generator of G is then a primitive nth root of unity.

To describe cyclotomic extensions, we need to use the *Euler phi function*. If n is a positive integer, let $\phi(n)$ be the number of integers between 1 and n that are relatively prime to n. The problems below give the main properties of the Euler phi function. We also need to know about the group of units of the ring $\mathbb{Z}/n\mathbb{Z}$. Recall that if R is a commutative ring with 1, then the set

$$R^* = \{a \in R : \text{there is a } b \in R \text{ with } ab = 1\}$$

is a group under multiplication; it is called the group of units of R. If $R = \mathbb{Z}/n\mathbb{Z}$, then an easy exercise shows that

$$(\mathbb{Z}/n\mathbb{Z})^* = \{a + n\mathbb{Z} : \gcd(a, n) = 1\}.$$

Therefore, $|(\mathbb{Z}/n\mathbb{Z})^*| = \phi(n)$.

We now describe cyclotomic extensions of an arbitrary base field.

Proposition 7.2 *Suppose that* $\operatorname{char}(F)$ *does not divide* n*, and let* K *be a splitting field of* $x^n - 1$ *over* F*. Then* K/F *is Galois,* $K = F(\omega)$ *is generated by any primitive nth root of unity* ω*, and* $\operatorname{Gal}(K/F)$ *is isomorphic to a subgroup of* $(\mathbb{Z}/n\mathbb{Z})^*$*. Thus,* $\operatorname{Gal}(K/F)$ *is Abelian and* $[K : F]$ *divides* $\phi(n)$*.*

Proof. Since $\operatorname{char}(F)$ does not divide n, the derivative test shows that $x^n - 1$ is a separable polynomial over F. Therefore, K is both normal and separable over F; hence, K is Galois over F. Let $\omega \in K$ be a primitive nth root of unity. Then all nth roots of unity are powers of ω, so $x^n - 1$ splits over $F(\omega)$. This proves that $K = F(\omega)$. Any automorphism of K that fixes F is determined by what it does to ω. However, any automorphism

restricts to a group automorphism of the set of roots of unity, so it maps the set of primitive nth roots of unity to itself. Any primitive nth root of unity in K is of the form ω^t for some t relatively prime to n. Therefore, the map $\theta : \mathrm{Gal}(K/F) \longrightarrow (\mathbb{Z}/n\mathbb{Z})^*$ given by $\sigma \mapsto t + n\mathbb{Z}$, where $\sigma(\omega) = \omega^t$, is well defined. If $\sigma, \tau \in \mathrm{Gal}(K/F)$ with $\sigma(\omega) = \omega^t$ and $\tau(\omega) = \omega^s$, then $(\sigma\tau)(\omega) = \sigma(\omega^s) = \omega^{st}$, so θ is a group homomorphism. The kernel of θ is the set of all σ with $\sigma(\omega) = \omega$; that is, $\ker(\theta) = \langle \mathrm{id} \rangle$. Thus, θ is injective, so $\mathrm{Gal}(K/F)$ is isomorphic to a subgroup of the Abelian group $(\mathbb{Z}/n\mathbb{Z})^*$, a group of order $\phi(n)$. This finishes the proof. □

Example 7.3 The structure of F determines the degree $[F(\omega) : F]$ or, equivalently, the size of $\mathrm{Gal}(F(\omega)/F)$. For instance, let $\omega = e^{2\pi i/8}$ be a primitive eighth root of unity in $\mathbb{C}$. Then $\omega^2 = i$ is a primitive fourth root of unity. The degree of $\mathbb{Q}(\omega)$ over $\mathbb{Q}$ is 4, which we will show below. If $F = \mathbb{Q}(i)$, then the degree of $F(\omega)$ over F is 2, since ω satisfies the polynomial $x^2 - i$ over F and $\omega \notin F$. If $F = \mathbb{R}$, then $\mathbb{R}(\omega) = \mathbb{C}$, so $[\mathbb{R}(\omega) : \mathbb{R}] = 2$. In fact, if $n \geq 3$ and if τ is any primitive nth root of unity in $\mathbb{C}$, then $\mathbb{R}(\tau) = \mathbb{C}$, so $[\mathbb{R}(\tau) : \mathbb{R}] = 2$.

Example 7.4 Let $F = \mathbb{F}_2$. If ω is a primitive third root of unity over F, then ω is a root of $x^3 - 1 = (x-1)(x^2+x+1)$. Since $\omega \neq 1$ and x^2+x+1 is irreducible over F, we have $[F(\omega) : F] = 2$ and $\min(F, \omega) = x^2 + x + 1$. If ρ is a primitive seventh root of unity, then by factoring $x^7 - 1$, by trial and error or by computer, we get

$$x^7 - 1 = (x-1)\left(x^3 + x + 1\right)\left(x^3 + x^2 + 1\right).$$

The minimal polynomial of ω is then one of these cubics, so $[F(\omega) : F] = 3$. Of the six primitive seventh roots of unity, three have $x^3 + x + 1$ as their minimal polynomial, and the three others have $x^3 + x^2 + 1$ as theirs. This behavior is different from cyclotomic extensions of $\mathbb{Q}$, as we shall see below, since all the primitive nth roots of unity over $\mathbb{Q}$ have the same minimal polynomial.

We now investigate cyclotomic extensions of $\mathbb{Q}$. Let $\omega_1, \ldots, \omega_r$ be the primitive nth roots of unity in $\mathbb{C}$. Then

$$\{\omega_1, \ldots, \omega_r\} = \left\{e^{2\pi i r/n} : \gcd(r, n) = 1\right\},$$

so there are $\phi(n)$ primitive nth roots of unity in $\mathbb{C}$. In Theorem 7.7, we will determine the minimal polynomial of a primitive nth root of unity over $\mathbb{Q}$, and so we will determine the degree of a cyclotomic extension of $\mathbb{Q}$.

Definition 7.5 *The nth cyclotomic polynomial is $\Psi_n(x) = \prod_{i=1}^{r}(x - \omega_i)$, the monic polynomial in $\mathbb{C}[x]$ whose roots are exactly the primitive nth roots of unity in $\mathbb{C}$.*

For example,

$$\begin{aligned}\Psi_1(x) &= x - 1,\\ \Psi_2(x) &= x + 1,\\ \Psi_4(x) &= (x - i)(x + i) = x^2 + 1.\end{aligned}$$

Moreover, if p is prime, then all pth roots of unity are primitive except for the root 1. Therefore,

$$\Psi_p(x) = (x^p - 1)/(x - 1) = x^{p-1} + x^{p-2} + \cdots + x + 1.$$

From this definition of $\Psi_n(x)$, it is not clear that $\Psi_n(x) \in \mathbb{Q}[x]$, nor that $\Psi_n(x)$ is irreducible over $\mathbb{Q}$. However, we verify the first of these facts in the next lemma and then the second in Theorem 7.7, which shows that $\Psi_n(x)$ is the minimal polynomial of a primitive nth root of unity over $\mathbb{Q}$.

Lemma 7.6 *Let n be any positive integer. Then $x^n - 1 = \prod_{d|n} \Psi_d(x)$. Moreover, $\Psi_n(x) \in \mathbb{Z}[x]$.*

Proof. We know that $x^n - 1 = \prod(x - \omega)$, where ω ranges over the set of all nth roots of unity. If d is the order of ω in $\mathbb{C}^*$, then d divides n, and ω is a primitive dth root of unity. Gathering all the dth root of unity terms together in this factorization proves the first statement. For the second, we use induction on n; the case $n = 1$ is clear since $\Psi_1(x) = x - 1$. Suppose that $\Psi_d(x) \in \mathbb{Z}[x]$ for all $d < n$. Then from the first part, we have

$$x^n - 1 = \left(\prod_{d|n,d<n} \Psi_d(x)\right) \cdot \Psi_n(x).$$

Since $x^n - 1$ and $\prod_{d|n} \Psi_d(x)$ are monic polynomials in $\mathbb{Z}[x]$, the division algorithm, Theorem 3.2 of Appendix A, shows that $\Psi_n(x) \in \mathbb{Z}[x]$. □

We can use this lemma to calculate the cyclotomic polynomials $\Psi_n(x)$ by recursion. For example, to calculate $\Psi_8(x)$, we have

$$x^8 - 1 = \Psi_8(x)\Psi_4(x)\Psi_2(x)\Psi_1(x),$$

so

$$\Psi_8(x) = \frac{x^8 - 1}{(x - 1)(x + 1)(x^2 + 1)} = x^4 + 1.$$

The next theorem is the main fact about cyclotomic polynomials and allows us to determine the degree of a cyclotomic extension over $\mathbb{Q}$.

Theorem 7.7 *Let n be any positive integer. Then $\Psi_n(x)$ is irreducible over $\mathbb{Q}$.*

Proof. To prove that $\Psi_n(x)$ is irreducible over $\mathbb{Q}$, suppose not. Since $\Psi_n(x) \in \mathbb{Z}[x]$ and is monic, $\Psi_n(x)$ is reducible over $\mathbb{Z}$ by Gauss' lemma. Say $\Psi_n = f(x)h(x)$ with $f(x), h(x) \in \mathbb{Z}[x]$ both monic and f irreducible over $\mathbb{Z}$. Let ω be a root of f. We claim that ω^p is a root of f for all primes p that do not divide n. If this is false for a prime p, then since ω^p is a primitive nth root of unity, ω^p is a root of h. Since $f(x)$ is monic, the division algorithm shows that $f(x)$ divides $h(x^p)$ in $\mathbb{Z}[x]$. The map $\mathbb{Z}[x] \to \mathbb{F}_p[x]$ given by reducing coefficients mod p is a ring homomorphism. For $g \in \mathbb{Z}[x]$, let $\overline{g}$ be the image of $g(x)$ in $\mathbb{F}_p[x]$. Reducing mod p yields $\overline{\Psi_n(x)} = \overline{f} \cdot \overline{h}$. Since $\overline{\Psi_n(x)}$ divides $x^n - \overline{1}$, the derivative test shows that $\overline{\Psi_n(x)}$ has no repeated roots in any extension field of $\mathbb{F}_p$, since p does not divide n. Now, since $a^p = a$ for all $a \in \mathbb{F}_p$, we see that $\overline{h(x^p)} = \overline{h(x)^p}$. Therefore, $\overline{f}$ divides $\overline{h^p}$, so any irreducible factor $\overline{q} \in \mathbb{F}_p[x]$ of $\overline{f}$ also divides $\overline{h}$. Thus, $\overline{q^2}$ divides $\overline{fh} = \overline{\Psi_n(x)}$, which contradicts the fact that $\overline{\Psi_n}$ has no repeated roots. This proves that if ω is a root of f, then ω^p is also a root of f, where p is a prime not dividing n. But this means that all primitive nth roots of unity are roots of f, for if α is a primitive nth root of unity, then $\alpha = \omega^t$ with t relatively prime to n. Then $\alpha = \omega^{p_1 \cdots p_r}$, with each p_i a prime relatively prime to n. We see that ω^{p_1} is a root of f, so then $(\omega^{p_1})^{p_2} = \omega^{p_1 p_2}$ is also a root of f. Continuing this shows α is a root of f. Therefore, every primitive nth root of unity is a root of f, so $\Psi_n(x) = f$. This proves that $\Psi_n(x)$ is irreducible over $\mathbb{Z}$, and so $\Psi_n(x)$ is also irreducible over $\mathbb{Q}$. □

If ω is a primitive nth root of unity in $\mathbb{C}$, then the theorem above shows that $\Psi_n(x)$ is the minimal polynomial of ω over $\mathbb{Q}$. The following corollary describes cyclotomic extensions of $\mathbb{Q}$.

Corollary 7.8 *If K is a splitting field of $x^n - 1$ over $\mathbb{Q}$, then $[K : \mathbb{Q}] = \phi(n)$ and $\mathrm{Gal}(K/\mathbb{Q}) \cong (\mathbb{Z}/n\mathbb{Z})^*$. Moreover, if ω is a primitive nth root of unity in K, then $\mathrm{Gal}(K/\mathbb{Q}) = \{\sigma_i : \gcd(i, n) = 1\}$, where σ_i is determined by $\sigma_i(\omega) = \omega^i$.*

Proof. The first part of the corollary follows immediately from Proposition 7.2 and Theorem 7.7. The description of $\mathrm{Gal}(K/\mathbb{Q})$ is a consequence of the proof of Proposition 7.2. □

If ω is a primitive nth root of unity in $\mathbb{C}$, then we will refer to the cyclotomic extension $\mathbb{Q}(\omega)$ as $\mathbb{Q}_n$.

Example 7.9 Let $K = \mathbb{Q}_7$, and let ω be a primitive seventh root of unity in $\mathbb{C}$. By Corollary 7.8, $\mathrm{Gal}(K/\mathbb{Q}) \cong (\mathbb{Z}/7\mathbb{Z})^*$, which is a cyclic group of order 6. The Galois group of $K/\mathbb{Q}$ is $\{\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6\}$, where $\sigma_i(\omega) = \omega^i$. Thus, $\sigma_1 = \mathrm{id}$, and it is easy to check that σ_3 generates this group. Moreover, $\sigma_i \circ \sigma_j = \sigma_{ij}$, where the subscripts are multiplied modulo 7. The

subgroups of $\mathrm{Gal}(K/\mathbb{Q})$ are then

$$\langle \mathrm{id} \rangle, \ \langle \sigma_3^3 \rangle, \ \langle \sigma_3^2 \rangle, \ \langle \sigma_3 \rangle,$$

whose orders are 1, 2, 3, and 6, respectively. Let us find the corresponding intermediate fields. If $L = \mathcal{F}(\sigma_3^3) = \mathcal{F}(\sigma_6)$, then $[K : L] = |\langle \sigma_6 \rangle| = 2$ by the fundamental theorem. To find L, we note that ω must satisfy a quadratic over L and that this quadratic is

$$(x - \omega)(x - \sigma_6(\omega)) = (x - \omega)(x - \omega^6).$$

Expanding, this polynomial is

$$x^2 - (\omega + \omega^6)x + \omega\omega^6 = x^2 - (\omega + \omega^6)x + 1.$$

Therefore, $\omega + \omega^6 \in L$. If we let $\omega = \exp(2\pi i/7) = \cos(2\pi/7) + i\sin(2\pi/7)$, then $\omega + \omega^6 = 2\cos(2\pi/7)$. Therefore, ω satisfies a quadratic over $\mathbb{Q}(\cos(2\pi/7))$; hence, L has degree at most 2 over this field. This forces $L = \mathbb{Q}(\cos(2\pi/7))$. With similar calculations, we can find $M = \mathcal{F}(\sigma_3^2) = \mathcal{F}(\sigma_2)$. The order of σ_2 is 3, so $[M : \mathbb{Q}] = 2$. Hence, it suffices to find one element of M that is not in $\mathbb{Q}$ in order to generate M. Let

$$\alpha = \omega + \sigma_2(\omega) + \sigma_2^2(\omega) = \omega + \omega^2 + \omega^4.$$

This element is in M because it is fixed by σ. But, we show that α is not in $\mathbb{Q}$ since it is not fixed by σ_6. To see this, we have

$$\begin{aligned} \sigma_6(\omega) &= \omega^6 + \omega^{12} + \omega^{24} \\ &= \omega^6 + \omega^5 + \omega^3. \end{aligned}$$

If $\sigma_6(\alpha) = \alpha$, this equation would give a degree 6 polynomial for which ω is a root, and this polynomial is not divisible by

$$\min(\mathbb{Q}, \omega) = \Psi_7(x) = x^6 + x^5 + x^4 + x^3 + x^2 + x + 1,$$

a contradiction. This forces $\alpha \notin \mathbb{Q}$, so $M = \mathbb{Q}(\alpha)$. Therefore, the intermediate fields of $K/\mathbb{Q}$ are

$$K, \ \mathbb{Q}(\cos(2\pi/7)), \ \mathbb{Q}(\omega + \omega^2 + \omega^4), \ \mathbb{Q}.$$

Example 7.10 Let $K = \mathbb{Q}_8$, and let $\omega = \exp(2\pi i/8) = (1 + i)/\sqrt{2}$. The Galois group of $K/\mathbb{Q}$ is $\{\sigma_1, \sigma_3, \sigma_5, \sigma_7\}$, and note that each of the three nonidentity automorphisms of K have order 2. The subgroups of this Galois group are then

$$\langle \mathrm{id} \rangle, \ \langle \sigma_3 \rangle, \ \langle \sigma_5 \rangle, \ \langle \sigma_7 \rangle, \ \mathrm{Gal}(K/\mathbb{Q}).$$

Each of the three proper intermediate fields has degree 2 over $\mathbb{Q}$. One is easy to find, since $\omega^2 = i$ is a primitive fourth root of unity. The group

associated to $\mathbb{Q}(i)$ is $\langle \sigma_5 \rangle$, since $\sigma_5(\omega^2) = \omega^{10} = \omega^2$. We could find the two other fields in the same manner as in the previous example: Show that the fixed field of σ_3 is generated over $\mathbb{Q}$ by $\omega + \sigma_3(\omega)$. However, we can get this more easily due to the special form of ω. Since $\omega = (1+i)/\sqrt{2}$ and $\omega^{-1} = (1-i)/\sqrt{2}$, we see that $\sqrt{2} = \omega + \omega^{-1} \in K$. The element $\omega + \omega^{-1} = \omega + \omega^7$ is fixed by σ_7; hence, the fixed field of σ_7 is $\mathbb{Q}(\sqrt{2})$. We know $i \in K$ and $\sqrt{2} \in K$, so $\sqrt{-2} \in K$. This element must generate the fixed field of σ_3. The intermediate fields are then

$$K,\ \mathbb{Q}(\sqrt{-2}),\ \mathbb{Q}(\sqrt{-1}),\ \mathbb{Q}(\sqrt{2}),\ \mathbb{Q}.$$

The description of the intermediate fields also shows that $K = \mathbb{Q}(\sqrt{2}, i)$.

Problems

1. Determine all of the subfields of $\mathbb{Q}_{12}$.

2. Show that $\cos(\pi/9)$ is algebraic over $\mathbb{Q}$, and find $[\mathbb{Q}(\cos(\pi/9)) : \mathbb{Q}]$.

3. Show that $\cos(2\pi/n)$ and $\sin(2\pi/n)$ are algebraic over $\mathbb{Q}$ for any $n \in \mathbb{N}$.

4. Prove that $\mathbb{Q}(\cos(2\pi/n))$ is Galois over $\mathbb{Q}$ for any n. Is the same true for $\mathbb{Q}(\sin(2\pi/n))$?

5. If p is a prime, prove that $\phi(p^n) = p^{n-1}(p-1)$.

6. Let $\theta : \mathbb{Z}[x] \to \mathbb{F}_p[x]$ be the map that sends $\sum_i a_i x^i$ to $\sum_i \overline{a_i} x^i$, where $\overline{a}$ is the equivalence class of a modulo p. Show that θ is a ring homomorphism.

7. If $\gcd(n, m) = 1$, show that $\phi(nm) = \phi(n)\phi(m)$.

8. If the prime factorization of n is $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$, show that $\phi(n) = \Pi_i p_i^{\alpha_i - 1}(p_i - 1)$.

9. Let n, m be positive integers with $d = \gcd(n, m)$ and $l = \operatorname{lcm}(n, m)$. Prove that $\phi(n)\phi(m) = \phi(d)\phi(l)$.

10. Show that $(\mathbb{Z}/n\mathbb{Z})^* = \{a + n\mathbb{Z} : \gcd(a, n) = 1\}$.

11. If n is odd, prove that $\mathbb{Q}_{2n} = \mathbb{Q}_n$.

12. Let n, m be positive integers with $d = \gcd(n, m)$ and $l = \operatorname{lcm}(n, m)$.
    (a) If n divides m, prove that $\mathbb{Q}_n \subseteq \mathbb{Q}_m$.
    (b) Prove that $\mathbb{Q}_n \mathbb{Q}_m = \mathbb{Q}_l$.
    (c) Prove that $\mathbb{Q}_n \cap \mathbb{Q}_m = \mathbb{Q}_d$.

13. Determine for which n and m there is an inclusion $\mathbb{Q}_n \subseteq \mathbb{Q}_m$. From this, determine which cyclotomic extensions contain $\sqrt{-1}$.

14. Find a positive integer n such that there is a subfield of $\mathbb{Q}_n$ that is not a cyclotomic extension of $\mathbb{Q}$.

15. If $d \in \mathbb{Q}$, show that $\mathbb{Q}(\sqrt{d})$ lies in some cyclotomic extension of $\mathbb{Q}$. (This is a special case of the Kronecker–Weber theorem, which states that any Galois extension of $\mathbb{Q}$ with Abelian Galois group lies in a cyclotomic extension of $\mathbb{Q}$.)

16. The group $(\mathbb{Z}/n\mathbb{Z})^*$ is a finite Abelian group; hence, it decomposes into a direct product of cyclic groups. This problem explicitly describes this decomposition.

    (a) If $n = p_1^{r_1} \cdots p_k^{r_k}$ is the prime factorization of n, show that $(\mathbb{Z}/n\mathbb{Z}) \cong \prod_i (\mathbb{Z}/p_i^{r_i}\mathbb{Z})$ as rings; hence, $(\mathbb{Z}/n\mathbb{Z})^* \cong \prod_i (\mathbb{Z}/p_i^{r_i}\mathbb{Z})^*$.

    (b) If p is an odd prime, show that $(1+p)^{p^t} \equiv 1 + p^{t+1} (\operatorname{mod} p^{t+2})$ if $t \geq 0$. Use this to find an element of large order in $\mathbb{Z}/p_i^{r_i}\mathbb{Z}$, and then conclude that $(\mathbb{Z}/p_i^{r_i}\mathbb{Z})^*$ is cyclic if p is an odd prime.

    (c) Show that $5^{2^t} \equiv 1 + 2^{t+2} (\operatorname{mod} 2^{t+3})$, and then that $(\mathbb{Z}/2^{r_i}\mathbb{Z})^* \cong \mathbb{Z}/2^{r-2}\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ if $r \geq 3$. Note that $(\mathbb{Z}/2^r\mathbb{Z})^*$ is cyclic if $r \leq 2$.

17. Let G be a finite Abelian group. Show that there is a Galois extension $K/\mathbb{Q}$ with $\operatorname{Gal}(K/\mathbb{Q}) \cong G$.

8 Norms and Traces

In this section, we define the norm and trace of a finite extension of fields and prove their basic properties. To help motivate these concepts, in Examples 7.9 and 7.10 we used elements of the form $\sum_{\sigma \in H} \sigma(\omega)$ to generate the intermediate field $\mathcal{F}(H)$ of a cyclotomic extension. We will see that the sum $\sum_{\sigma \in H} \sigma(\omega)$ is the trace of ω in the extension $K/\mathcal{F}(H)$. The definitions we give will not look like these sums; instead, we define the norm and trace in terms of linear transformations. This approach generalizes more readily to other situations. For instance, given a division ring (finite dimensional over its center), there is a notion of norm and trace that is quite important.

Let K be a field extension of F with $[K : F] = n$. If $a \in K$, let L_a be the map $L_a : K \longrightarrow K$ given by $L_a(b) = ab$. It is easy to see that L_a is an F-vector space homomorphism. Since K is a finite dimensional F-vector space, we can view F-linear transformations of K as matrices by using bases; that is, if $\operatorname{End}_F(K) = \hom_F(K, K)$ is the ring of all F-vector space homomorphisms from K to K, then there is an isomorphism $\operatorname{End}_F(K) \cong M_n(F)$, where $M_n(F)$ is the ring of $n \times n$ matrices over F.

If $\varphi : \mathrm{End}_F(K) \to M_n(F)$ is an isomorphism, we can use φ to define the determinant and trace of a linear transformation. If $T \in \mathrm{End}_F(K)$, let $\det(T) = \det(\varphi(T))$ and $\mathrm{Tr}(T) = \mathrm{Tr}(\varphi(T))$. These definitions do not depend on φ; to see this, let ψ be another isomorphism. Then ψ corresponds to choosing a basis for K different from that used to obtain φ. Therefore, the two matrix representations of a transformation T are similar; that is, there is an invertible matrix A with $\psi(T) = A^{-1}\varphi(T)A$. Therefore, $\det(\psi(T)) = \det(\varphi(T))$ and $\mathrm{Tr}(\psi(T)) = \mathrm{Tr}(\varphi(T))$.

Definition 8.1 *Let K be a finite extension of F. The norm $N_{K/F}$ and trace $T_{K/F}$ are defined for all $a \in K$ by*

$$N_{K/F}(a) = \det(L_a),$$
$$T_{K/F}(a) = \mathrm{Tr}(L_a).$$

Example 8.2 Let F be any field, and let $K = F(\sqrt{d})$ for some $d \in F - F^2$. A convenient basis for K is $\{1, \sqrt{d}\}$. If $\alpha = a + b\sqrt{d}$ with $a, b \in F$, we determine the norm and trace of α. The linear transformation L_α is equal to $aL_1 + bL_{\sqrt{d}}$, so we first need to find the matrix representations for L_1 and $L_{\sqrt{d}}$. The identity transformation L_1 has matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. For $L_{\sqrt{d}}$, we have

$$L_{\sqrt{d}}(1) = \sqrt{d} = 0 \cdot 1 + 1 \cdot \sqrt{d},$$
$$L_{\sqrt{d}}(\sqrt{d}) = d = d \cdot 1 + 0 \cdot \sqrt{d}.$$

Therefore, the matrix for $L_{\sqrt{d}}$ is $\begin{pmatrix} 0 & d \\ 1 & 0 \end{pmatrix}$. The matrix for L_α is then

$$a\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + b\begin{pmatrix} 0 & d \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a & bd \\ b & a \end{pmatrix}.$$

From this we obtain $N_{K/F}(a + b\sqrt{d}) = a^2 - b^2 d$ and $T_{K/F}(a + b\sqrt{d}) = 2a$. In particular, $N_{K/F}(\sqrt{d}) = -d$ and $T_{K/F}(\sqrt{d}) = 0$.

Example 8.3 Let $F = \mathbb{Q}$ and $K = \mathbb{Q}(\sqrt[3]{2})$. We will determine the norm and trace of $\sqrt[3]{2}$. An F-basis for K is $\{1, \sqrt[3]{2}, \sqrt[3]{4}\}$. We can check that $L_{\sqrt[3]{2}}(1) = \sqrt[3]{2}$, $L_{\sqrt[3]{2}}(\sqrt[3]{2}) = \sqrt[3]{4}$, and $L_{\sqrt[3]{2}}(\sqrt[3]{4}) = 2$. Therefore, the matrix representing $L_{\sqrt[3]{2}}$ using this basis is

$$\begin{pmatrix} 0 & 0 & 2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

so $N_{K/F}(\sqrt[3]{2}) = 2$ and $T_{K/F}(\sqrt[3]{2}) = 0$.

Example 8.4 Let F be a field of characteristic $p > 0$, and let K/F be a purely inseparable extension of degree p. Say $K = F(\alpha)$ with $\alpha^p = a \in F$. For instance, we could take K to be the rational function field $k(x)$ over a field k of characteristic p and $F = k(x^p)$. A basis for K is $\{1, \alpha, \alpha^2, \ldots, \alpha^{p-1}\}$. With respect to this basis, the matrix for L_α is

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & a \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & & 1 & 0 \end{pmatrix}.$$

We leave it to the reader to check that the matrix for L_{α^2} is obtained by taking this matrix and shifting the columns to the left, moving the first column to the end. Similar processes yield the matrices L_{α^i} for each i. From these matrices, we see that $N_{K/F}(\alpha) = (-1)^p a$. For traces, each L_{α^i} has trace 0, including the identity matrix, since $p \cdot 1 = 0$ in F. Therefore, for any $\beta \in K$ we have $\mathrm{Tr}_{K/F}(\beta) = 0$. The trace map $T_{K/F}$ is thus the zero function.

The following lemma gives some elementary properties of norm and trace.

Lemma 8.5 *Let K be a finite extension of F with $n = [K : F]$.*

1. *If $a \in K$, then $N_{K/F}(a)$ and $T_{K/F}(a)$ lie in F.*
2. *The trace map $T_{K/F}$ is an F-linear transformation.*
3. *If $\alpha \in F$, then $T_{K/F}(\alpha) = n\alpha$.*
4. *If $a, b \in K$, then $N_{K/F}(ab) = N_{K/F}(a) \cdot N_{K/F}(b)$.*
5. *If $\alpha \in F$, then $N_{K/F}(\alpha) = \alpha^n$.*

Proof. These properties all follow immediately from the definitions and properties of the determinant and trace functions. □

The examples above indicate that it is not easy in general to calculate norms and traces from our definition. In order to work effectively with norms and traces, we need alternative ways of calculating them. The next proposition shows that if we know the minimal polynomial of an element, then it is easy to determine the norm and trace of that element.

Proposition 8.6 *Let K be an extension of F with $[K : F] = n$. If $a \in K$ and $p(x) = x^m + \alpha_{m-1}x^{m-1} + \cdots + \alpha_1 x + \alpha_0$ is the minimal polynomial of a over F, then $N_{K/F}(a) = (-1)^n \alpha_0^{n/m}$ and $T_{K/F}(a) = -\frac{n}{m}\alpha_{m-1}$.*

Proof. Let $\varphi : K \longrightarrow \mathrm{End}_F(K)$ be the map $\varphi(a) = L_a$. It is easy to see that $L_{a+b} = L_a + L_b$ and $L_{ab} = L_a \circ L_b$, so φ is a ring homomorphism. Also, if $\alpha \in F$ and $a \in K$, then $L_{\alpha a} = \alpha L_a$. Thus, φ is also an F-vector space homomorphism. The kernel of φ is necessarily trivial, since φ is not the zero map. Since φ is injective, the minimal polynomials of a and L_a are equal. Let $\chi(x)$ be the characteristic polynomial of L_a, and say $\chi(x) = x^n + \beta_{n-1}x^{n-1} + \cdots + \beta_0$. By the Cayley–Hamilton theorem, Theorem 2.1 of Appendix D, the characteristic and minimal polynomials of a linear transformation have the same irreducible factors, and the minimal polynomial divides the characteristic polynomial. Since p is irreducible, by comparing degrees we see that $\chi(x) = p(x)^{n/m}$. Note that m divides n, because $m = [F(a) : F]$ and

$$n = [K : F] = [K : F(a)] \cdot [F(a) : F].$$

Now, recalling the relation between the determinant and trace of a matrix and its characteristic polynomial, we see that $N_{K/F}(a) = \det(L_a) = (-1)^n\beta_0$ and $T_{K/F}(a) = \mathrm{Tr}(L_a) = -\beta_{n-1}$. Multiplying out $p(x)^{n/m}$ shows that $\beta_0 = \alpha_0^{n/m}$ and $\beta_{n-1} = \frac{n}{m}\alpha_{m-1}$, which proves the proposition. □

Example 8.7 If F is any field and if $K = F(\sqrt{d})$ for some $d \in F - F^2$, then a short calculation shows that the minimal polynomial of $a + b\sqrt{d}$ is $x^2 - 2ax + (a^2 - b^2d)$. Proposition 8.6 yields $N_{K/F}(a + b\sqrt{d}) = a^2 - b^2d$ and $T_{K/F}(a + b\sqrt{d}) = 2a$, as we had obtained before.

If $F = \mathbb{Q}$ and $K = \mathbb{Q}(\sqrt[3]{2})$, then the minimal polynomial of $\sqrt[3]{2}$ over F is $x^3 - 2$. Then $N_{K/F}(\sqrt[3]{2}) = 2$ and $T_{K/F}(\sqrt[3]{2}) = 0$.

Example 8.8 If K is a purely inseparable extension of F of characteristic p, then the minimal polynomial of any element of K is of the form $x^{p^n} - a$. From this, it follows that the trace of any element is zero.

If we know the minimal polynomial of an element, then it is easy to find the norm and trace of the element. However, it may be hard to find the minimal polynomial in many situations. Therefore, additional methods of calculating norms and traces are needed. For a Galois extension K of F, there are simple descriptions of norm and trace in terms of automorphisms. Theorem 8.12 describes the norm and trace in terms of F-homomorphisms for general finite extensions and has the description for Galois extensions as a special case. In order to prove this result, we need some facts about separable and purely inseparable closures. Let K be a finite extension of F, and let S be the separable closure of F in K. Recall that the *purely inseparable degree* of K/F is $[K : F]_i = [K : S]$. The next three lemmas prove the facts we need in order to obtain the descriptions of norms and traces that we desire.

Lemma 8.9 *Let K be a finite extension of F, and let S be the separable closure of F in K. Then $[S : F]$ is equal to the number of F-homomorphisms from K to an algebraic closure of F.*

Proof. Let M be an algebraic closure of F. We may assume that $K \subseteq M$. If S is the separable closure of F in K, then $S = F(a)$ for some a by the primitive element theorem. If $r = [S : F]$, then there are r distinct roots of $\min(F, a)$ in M. Suppose that these roots are $a_1, \ldots, a_r$. Then the map $\sigma_i : S \to M$ defined by $f(a) = f(a_i)$ is a well-defined F-homomorphism for each i. Moreover, any F-homomorphism from S to M must be of this form since a must map to a root of $\min(F, a)$. Therefore, there are r distinct F-homomorphisms from S to M. The field K is purely inseparable over S; hence, K is normal over S. Therefore, each σ_i extends to an F-homomorphism from K to M by Proposition 3.28. We will be done once we show that each σ_i extends in a unique way to K. To prove this, suppose that τ and ρ are extensions of σ_i to K. Then $\tau(K) = K$ by Proposition 3.28, and so $\tau^{-1}\rho$ is an automorphism of K that fixes S. However, $\mathrm{Gal}(K/S) = \{\mathrm{id}\}$, since K/S is purely inseparable. Therefore, $\tau^{-1}\rho = \mathrm{id}$, so $\tau = \rho$. □

Lemma 8.10 *Let K be a finite dimensional, purely inseparable extension of F. If $a \in K$, then $a^{[K:F]} \in F$. More generally, if N is a finite dimensional, Galois extension of F and if $a \in NK$, then $a^{[K:F]} \in N$.*

Proof. Let K be purely inseparable over F, and let $n = [K : F]$. If $a \in K$, then $a^{[F(a):F]} \in F$ by Lemma 4.16. Since $[F(a) : F]$ divides $n = [K : F]$, we also have $a^n \in F$. To prove the second statement, let N be a Galois extension of F. Then $N \cap K$ is both separable and purely inseparable over F, so $N \cap K = F$. Therefore, $[NK : K] = [N : F]$ by the theorem of natural irrationalities, so $[NK : N] = [K : F]$. The extension NK/N is purely inseparable, so by the first part of the proof, we have $a^n \in N$ for all $a \in NK$. This finishes the proof. □

Lemma 8.11 *Suppose that $F \subseteq L \subseteq K$ are fields with $[K : F] < \infty$. Then $[K : F]_i = [K : L]_i \cdot [L : F]_i$.*

Proof. Let S_1 be the separable closure of F in L, let S_2 be the separable closure of L in K, and let S be the separable closure of F in K. Since any element of K that is separable over F is also separable over L, we see that $S \subseteq S_2$. Moreover, SL is a subfield of S_2 such that S_2/SL is both separable and purely inseparable, so $S_2 = SL$. We claim that this means that $[L : S_1] = [S_2 : S]$. If this is true, then

$$\begin{aligned}[K : F]_i &= [K : S] \\ &= [K : S_2] \cdot [S_2 : S] \\ &= [K : S_2] \cdot [L : S_1]\end{aligned}$$

$$= [K : L]_i \cdot [L : F]_i,$$

proving the result. We now verify that $[L : S_1] = [S_2 : S]$. By the primitive element theorem, $S = S_1(a)$ for some a. Let $f(x) = \min(S_1, a)$, and let $g(x) = \min(L, a)$. Then g divides f in $L[x]$. However, since L is purely inseparable over S_1, some power of g lies in $S_1[x]$. Consequently, f divides a power of g in $F[x]$. These two divisibilities force f to be a power of g. The polynomial f has no repeated roots since a is separable over S_1, so the only possibility is for $f = g$. Thus, $[S : S_1] = [L(a) : L]$, and since $L(a) = SL = S_2$, we see that $[S : S_1] = [S_2 : L]$. Therefore,

$$[S_2 : S] = \frac{[S_2 : S_1]}{[S : S_1]} = \frac{[S_2 : S_1]}{[S_2 : L]} = [L : S_1].$$

This finishes the proof. □

We are now in the position to obtain the most useful description of the norm and trace of an element. The next theorem gives formulas that are particularly useful for a Galois extension and will allow us to prove a transitivity theorem for norms and traces.

Theorem 8.12 *Let K be a finite extension of F, and let $\sigma_1, \dots, \sigma_r$ be the distinct F-homomorphisms from K to an algebraic closure of F. If $a \in K$, then*

$$N_{K/F}(a) = \left(\prod_j \sigma_j(a)\right)^{[K:F]_i} \quad \textit{and} \quad T_{K/F}(a) = [K : F]_i \sum_j \sigma_j(a).$$

Proof. Let M be an algebraic closure of F, and let $\sigma_1, \dots, \sigma_r$ be the distinct F-homomorphisms from K to M. Let $g(x) = \left(\prod_j x - \sigma_j(a)\right)^{[K:F]_i}$, a polynomial over M. If S is the separable closure of F in K, then $r = [S : F]$ by Lemma 8.9. The degree of g is

$$\begin{aligned} r[K : F]_i &= r[K : S] = [K : S] \cdot [S : F] \\ &= [K : F] = n. \end{aligned}$$

We claim that $g(x) \in F[x]$ and that $g(x)$ has precisely the same roots as $p(x) = \min(F, a)$. If this is true, we see that p divides g, and since all roots of g are roots of p, the only irreducible factor of g is p. Thus, $g(x) = p(x)^{n/m}$, where $m = \deg(p(x))$. It was shown in the proof of Theorem 8.6 that $p^{n/m}$ is the characteristic polynomial $\chi(x)$ of L_a. Thus, $g(x) = \chi(x)$. Therefore, if $g(x) = x^n + \gamma_{n-1}x^{n-1} + \cdots + \gamma_0$, we have $N_{K/F}(a) = (-1)^n\gamma_0$ and $T_{K/F}(a) = -\gamma_{n-1}$. Multiplying out $g(x)$ shows that

$$\gamma_0 = \left(\prod_j -\sigma_j(a)\right)^{[K:F]_i}$$

and

$$\gamma_{n-1} = -[K : F]_i \sum_j \sigma_j(a).$$

The formulas for the norm and trace then follow from Proposition 8.6.

To see that $g(x) \in F[x]$ and that g and p have the same roots, first note that each $\sigma_j(a)$ is a root of p since σ_j is an F-homomorphism. If $b \in M$ is a root of $p(x)$, then by the isomorphism extension theorem there is a $\tau : M \longrightarrow M$ with $\tau(a) = b$. Since $\tau|_K$ is one of the σ_j, say $\tau|_K = \sigma_k$, then $\tau(a) = \sigma_k(a) = b$, so b is a root of g. This proves that g and p have the same roots. To see that $g(x) \in F[x]$, let N be the normal closure of S/F. Then N/F is Galois; hence, N/F is separable. Also, KN/K is Galois, and by the theorem of natural irrationalities, $[KN : K]$ divides $[N : S]$. Therefore, $[KN : N]$ divides $[K : S] = [K : F]_i$, since

$$[KN : N] \cdot [N : S] = [KN : S] = [KN : K] \cdot [K : S].$$

The extension KN/N is purely inseparable since K/S is purely inseparable, so $c^{[K:F]_i} \in N$ for any $c \in KN$ by Lemma 8.10. Because KN is the composite of a Galois extension of S with a purely inseparable, hence normal, extension, KN/S is normal. Thus, $\sigma_j(K) \subseteq KN$ by Proposition 3.28. So we see that $g(x) \in N[x]$, using $(KN)^{[K:F]_i} \subseteq N$. However, if τ is any element of $\mathrm{Gal}(M/N)$, then

$$\{(\tau\sigma_1)|_K, \ldots, (\tau\sigma_r)|_K\} = \{\sigma_1, \ldots, \sigma_r\},$$

so $\tau(g) = g$. Thus, the coefficients of g lie in the fixed field of $\mathrm{Gal}(M/F)$. This fixed field is the purely inseparable closure of F in M, since M/F is normal. We have seen that the coefficients of g lie in N, so they are separable over F. These coefficients must then be in F. This completes the proof of the theorem. □

Suppose that K is Galois over F. Then $\{\sigma_1, \ldots, \sigma_r\} = \mathrm{Gal}(K/F)$ and $[K : F]_i = 1$. The following corollary is immediate from Theorem 8.12.

Corollary 8.13 *If K/F is Galois with Galois group G, then for all $a \in K$,*

$$N_{K/F}(a) = \prod_{\sigma \in G} \sigma(a) \quad \textit{and} \quad T_{K/F}(a) = \sum_{\sigma \in G} \sigma(a).$$

Example 8.14 Let F be a field of characteristic not 2, and let $K = F(\sqrt{d})$ for some $d \in F - F^2$. Then $\mathrm{Gal}(K/F) = \{\mathrm{id}, \sigma\}$, where $\sigma(\sqrt{d}) = -\sqrt{d}$. Therefore,

$$N_{K/F}(a + b\sqrt{d}) = (a + b\sqrt{d})(a - b\sqrt{d}) = a^2 - b^2 d,$$
$$T_{K/F}(a + b\sqrt{d}) = (a + b\sqrt{d}) + (a - b\sqrt{d}) = 2a.$$

Example 8.15 Suppose that F is a field containing a primitive nth root of unity ω, and let K be an extension of F of degree n with $K = F(\alpha)$ and $\alpha^n = a \in F$. By the isomorphism extension theorem, there is an automorphism of K with $\sigma(\alpha) = \omega\alpha$. From this, we can see that the order of σ is n, so $\text{Gal}(K/F) = \langle\sigma\rangle$. Therefore,

$$\begin{aligned} N_{K/F}(\alpha) &= \alpha\sigma(\alpha)\cdots\sigma^{n-1}(\alpha) = \alpha \cdot \omega\alpha \cdots \omega^{n-1}\alpha \\ &= \omega^{n(n-1)/2}\alpha^n = (-1)^n a. \end{aligned}$$

If n is odd, then $n(n-1)/2$ is a multiple of n, so $\omega^{n(n-1)/2} = 1$. If n is even, then this exponent is not a multiple of n, so $\omega^{n(n-1)/2} \neq 1$. However, $(\omega^{n(n-1)/2})^2 = 1$, so $\omega^{n(n-1)/2} = -1$. This justifies the final equality $N_{K/F}(\alpha) = (-1)^n a$.

As for the trace,

$$\begin{aligned} T_{K/F}(\alpha) &= \alpha + \omega\alpha + \cdots + \omega^{n-1}\alpha = (1 + \omega + \cdots + \omega^{n-1})\alpha \\ &= 0 \end{aligned}$$

because ω is a root of $(x^n - 1)/(x - 1) = 1 + x + \cdots + x^{n-1}$. These norm and trace calculations could also have been obtained by using the minimal polynomial of α, which is $x^n - a$.

In the examples above, we often calculated the norm and trace of an element α for the field extension $F(\alpha)/F$. If we want the norm and trace of an element that does not generate the larger field, our calculations will be more involved. This complication is eliminated by the following transitivity theorem, which gives relations between the norm and trace of an extension and a subextension.

Theorem 8.16 *If $F \subseteq L \subseteq K$ are fields with $[K : F] < \infty$, then*

$$N_{K/F} = N_{L/F} \circ N_{K/L} \quad \textit{and} \quad T_{K/F} = T_{L/F} \circ T_{K/L};$$

that is, $N_{K/F}(a) = N_{L/F}(N_{K/L}(a))$ and $T_{K/F}(a) = T_{L/F}(T_{K/L}(a))$ for each $a \in K$.

Proof. Let M be an algebraic closure of F, let $\sigma_1, \ldots, \sigma_r$ be the distinct F-homomorphisms of L to M, and let $\tau_1, \ldots, \tau_s$ be the distinct L-homomorphisms of K to M. By the isomorphism extension theorem, we can extend each σ_j and τ_k to automorphisms $M \longrightarrow M$, which we will also call σ_j and τ_k, respectively. Each $\sigma_j\tau_k$ is an F-homomorphism from K to M. In fact, any F-homomorphism of K to M is of this type, as we now prove. If $\rho : K \longrightarrow M$ is an F-homomorphism, then $\rho|_L : L \longrightarrow M$ is equal to σ_j for some j. The map $\sigma_j^{-1}\rho$ is then an F-homomorphism $K \longrightarrow M$ which fixes L. Thus, $\sigma_j^{-1}\rho = \tau_k$ for some k, so $\rho = \sigma_j\tau_k$. If $a \in K$, then by

Theorem 8.12 we have

$$N_{K/F}(a) = \left(\prod_{j,k} \sigma_j \tau_k(a)\right)^{[K:F]_i} \quad \text{and} \quad N_{K/L}(a) = \left(\prod_k \tau_k(a)\right)^{[K:L]_i}.$$

Therefore,

$$\begin{aligned} N_{L/F}(N_{K/L}(a)) &= \left(\prod_j \sigma_j \left(\prod_k \tau_k(a)\right)^{[K:L]_i}\right)^{[L:F]_i} \\ &= \left(\prod_{j,k} \sigma_j \tau_k(a)\right)^{[K:L]_i [L:F]_i}. \end{aligned}$$

Since $[K : F]_i = [K : L]_i \cdot [L : F]_i$ by Lemma 8.11, this proves that $N_{K/F}(a) = N_{L/F}(N_{K/L}(a))$. A similar calculation shows that $T_{K/F}(a) = T_{L/F}(T_{K/L}(a))$. □

As a consequence of this theorem, we see in the following corollary that the existence of an element with nonzero trace is a test for separability.

Corollary 8.17 *A finite extension K/F is separable if and only if $T_{K/F}$ is not the zero map; that is, K/F is separable if and only if there is an $a \in K$ with $T_{K/F}(a) \neq 0$.*

Proof. Suppose that K/F is not separable. Then $\text{char}(F) = p > 0$. Let S be the separable closure of F in K. Then $S \neq K$ and K/S is a purely inseparable extension. Moreover, $[K : S] = p^t$ for some $t \geq 1$ by Lemma 4.17. If $a \in K$, then by Theorem 8.16 we have $T_{K/F}(a) = T_{S/F}(T_{K/S}(a))$. However by Theorem 8.12, if $\sigma_1, \ldots, \sigma_r$ are the distinct S-homomorphisms from K to an algebraic closure of F, then

$$T_{K/S}(a) = [K : S]_i(\sigma_1(a) + \cdots + \sigma_r(a)).$$

But $[K : S]_i = [K : S] = p^t$, since K is purely inseparable over S. Since $\text{char}(F) = p$, this forces $T_{K/S}(a) = 0$, so $T_{K/F}(a) = T_{S/F}(0) = 0$. Thus, $T_{K/F}$ is the zero map.

Conversely, suppose that K is separable over F. Let N be the normal closure of K/F. By Theorem 8.16, we see that if $T_{N/F}$ is nonzero, then so is $T_{K/F}$. Say $\text{Gal}(N/F) = \{\sigma_1, \ldots, \sigma_n\}$. If $a \in N$, then $T_{N/F}(a) = \sum_j \sigma_j(a)$ by the corollary to Theorem 8.12. By Dedekind's lemma, $\sigma_1(a) + \cdots + \sigma_n(a)$ is not zero for all $a \in N$, so $T_{N/F}$ is not the zero map. Therefore, $T_{K/F}$ is not the zero map. □

Problems

1. Let K/F be an extension of finite fields. Show that the norm map $N_{K/F}$ is surjective.

2. Let p be an odd prime, let ω be a primitive pth root of unity, and let $K = \mathbb{Q}(\omega)$. Show that $N_{K/\mathbb{Q}}(1 - \omega) = p$.

3. Let $n \geq 3$ be an integer, let ω be a primitive nth root of unity, and let $K = \mathbb{Q}(\omega)$. Show that $N_{K/\mathbb{Q}}(\omega) = 1$.

4. In Examples 7.9 and 7.10, generators consisting of traces were found for intermediate fields. Let K be a Galois extension of F. If L is an intermediate field of K/F, show that L is generated over F by traces from K to L. In other words, show that $L = F(\{T_{K/L}(a) : a \in K\})$.

5. Let K be a Galois extension of F. Prove or disprove that any intermediate field L of K/F is of the form $L = F(\{N_{K/L}(a) : a \in K\})$.

6. Let $F \subseteq K \subseteq L$ be fields with L/F a finite extension. Use the product theorem for the purely inseparable degree proved in this section to prove the corresponding product formula for separable degree; that is, prove that $[L : F]_s = [L : K]_s[K : F]_s$.

9 Cyclic Extensions

We resume our investigation of special types of Galois extensions. In this section, we study Galois extensions with cyclic Galois group. Section 11 will study Galois extensions with an Abelian Galois group.

Definition 9.1 *A Galois extension K/F is called cyclic if* $\mathrm{Gal}(K/F)$ *is a cyclic group.*

Example 9.2 Let F be a field of characteristic not 2, and let $a \in F^* - F^{*2}$. If $K = F(\sqrt{a})$, then $\mathrm{Gal}(K/F) = \{\mathrm{id}, \sigma\}$ where $\sigma(\sqrt{a}) = -\sqrt{a}$. Thus, $\mathrm{Gal}(K/F) \cong \mathbb{Z}/2\mathbb{Z}$ is cyclic. For another example, if p is a prime, then the cyclotomic extension $\mathbb{Q}_p/\mathbb{Q}$ is cyclic, since $\mathrm{Gal}(\mathbb{Q}_p/\mathbb{Q}) \cong (\mathbb{Z}/p\mathbb{Z})^*$ is a cyclic group.

Example 9.3 Let ω be a primitive fifth root of unity in $\mathbb{C}$, let $F = \mathbb{Q}(\omega)$, and let $K = F(\sqrt[5]{2})$. Then K is the splitting field of $x^5 - 2$ over F, so K is Galois over F. Also, $[F : \mathbb{Q}] = 4$ and $[\mathbb{Q}(\sqrt[5]{2}) : \mathbb{Q}] = 5$. The field K is the composite of these two extensions of $\mathbb{Q}$. The degree $[K : \mathbb{Q}]$ is divisible by 4 and 5; hence, it is divisible by 20. Moreover, $[K : F] \leq 5$, so $[K : \mathbb{Q}] \leq 20$. Therefore, $[K : \mathbb{Q}] = 20$, and so $[K : F] = 5$. Let $\alpha = \sqrt[5]{2}$. The roots of

$\min(F, \alpha)$ are α, $\omega\alpha$, $\omega^2\alpha$, $\omega^3\alpha$, and $\omega^4\alpha$. By the isomorphism extension theorem, there is a $\sigma \in \text{Gal}(K/F)$ with $\sigma(\alpha) = \omega\alpha$. Then $\sigma^i(\alpha) = \omega^i\alpha$. Consequently, $\sigma^5 = \text{id}$ and $\sigma^i \neq \text{id}$ if $i < 5$. The order of σ is thus equal to 5. This means that $\text{Gal}(K/F) = \langle\sigma\rangle$, so K/F is a cyclic extension.

We will analyze the cyclic extensions of degree n of a field containing a primitive nth root of unity and the cyclic extensions of degree p of a field of characteristic p. To motivate our restriction to these extensions, we first point out that there is no simple description of the cyclic extensions of degree n of a field F that does not contain a primitive nth root of unity, unless $n = p$ is a prime and $\text{char}(F) = p$. For instance, there is no simple description of the cyclic extensions of $\mathbb{Q}$, extensions that are important in algebraic number theory. Second, we can decompose a cyclic extension of F into a tower of degree p cyclic extensions together with a cyclic extension of degree relatively prime to p. We do this as follows. Let H be a p-Sylow subgroup of $G = \text{Gal}(K/F)$, and let $L = \mathcal{F}(H)$. Since H is normal in G, by the fundamental theorem L is Galois over F with $[L : F] = q$ and $[K : L] = p^n$. Furthermore, since subgroups and quotient groups of cyclic groups are cyclic, both L/F and K/L are cyclic extensions. Because $H = \text{Gal}(K/L)$ is a cyclic p-group, there is a chain of subgroups

$$\langle\text{id}\rangle \subset H_1 \subset H_2 \subset \cdots \subset H_n = H$$

with $|H_i| = p^i$. If $L_i = \mathcal{F}(H_i)$, we get a tower of fields

$$L_n = L \subset L_{n-1} \subset \cdots \subset L_0 = K.$$

Moreover, $[L_{m-1} : L_m] = p$ and L_{m-1} is a cyclic extension of L_m.

Let F be a field containing a primitive nth root of unity ω. If K is an extension of F, suppose that there exists an $a \in K$ with $a^n = b \in F$. We then write $a = \sqrt[n]{b}$. Note that $(\omega^i a)^n = b$ for all $i \in \mathbb{Z}$. Therefore, K contains n roots of the polynomial $x^n - b$, so $F(\sqrt[n]{b})$ is the splitting field of $x^n - b$ over F.

The following lemma is the heart of Theorem 9.5. The standard proof of this lemma is to use the Hilbert theorem 90. While we give a linear algebra proof of this, the Hilbert theorem 90 is quite important, and we discuss it in detail in Section 10.

Lemma 9.4 *Let F be a field containing a primitive nth root of unity ω, let K/F be a cyclic extension of degree n, and let σ be a generator of* $\text{Gal}(K/F)$. *Then there is an $a \in K$ with $\omega = \sigma(a)/a$.*

Proof. The automorphism σ is an F-linear transformation of K. We wish to find an $a \in K$ with $\sigma(a) = \omega a$; that is, we want to show that ω is an eigenvalue for σ. To do this, we show that ω is a root of the characteristic polynomial of σ. Now, since σ has order n in $\text{Gal}(K/F)$, we have $\sigma^n = \text{id}$.

Therefore, σ satisfies the polynomial $x^n - 1$. Moreover, if there is a polynomial $g(x) \in F[x]$ of degree $m < n$ satisfied by σ, then the automorphisms id, $\sigma, \ldots, \sigma^{m-1}$ are linearly dependent over F, a contradiction to the Dedekind independence lemma. Thus, $x^n - 1$ is the minimal polynomial of σ over F. However, the characteristic polynomial of σ has degree $n = [K : F]$ and is divisible by $x^n - 1$, so $x^n - 1$ is the characteristic polynomial of σ. Since ω is a root of this polynomial, ω is an eigenvalue for σ. Thus, there is an $a \in K$ with $\sigma(a) = \omega a$. □

We now give the description of cyclic extensions K/F of degree n when F contains a primitive nth root of unity.

Theorem 9.5 *Let F be a field containing a primitive nth root of unity, and let K/F be a cyclic Galois extension of degree n. Then there is an $a \in K$ with $K = F(a)$ and $a^n = b \in F$; that is, $K = F(\sqrt[n]{b})$.*

Proof. By the lemma, there is an a with $\sigma(a) = \omega a$. Therefore, $\sigma^i(a) = \omega^i a$, so a is fixed by σ^i only when n divides i. Since the order of σ is n, we see that a is fixed only by id, so $\mathrm{Gal}(K/F(a)) = \langle \mathrm{id} \rangle$. Thus, $K = F(a)$ by the fundamental theorem. We see that $\sigma(a^n) = (\omega a)^n = a^n$, so a^n is fixed by σ. Hence, $b = a^n \in F$, so $K = (\sqrt[n]{b})$. □

We give a converse to this theorem that describes extensions of the form $F(\sqrt[n]{b})/F$. This converse is a special case of a theorem we will see in Section 11.

Proposition 9.6 *Let F be a field containing a primitive nth root of unity, and let $K = F(\sqrt[n]{b})$ for some $b \in F$. Then K/F is a cyclic Galois extension. Moreover, $m = [K : F]$ is equal to the order of the coset bF^{*n} in the group F^*/F^{*n}, and $\min(F, \sqrt[n]{b}) = x^m - d$ for some $d \in F$.*

Proof. Let $a \in K$ with $a^n = b$. Since F contains a primitive nth root of unity ω, the polynomial $x^n - b$ splits over K, and it is separable over F by the derivative test. Thus, K is a splitting field over F for $x^n - b$, so K/F is Galois. We will show that K/F is cyclic Galois by determining a generator for $G = \mathrm{Gal}(K/F)$. The roots of $\min(F, a)$ lie in the set $\{\omega^j a : j \in \mathbb{Z}\}$ since $\min(F, a)$ divides $x^n - b$, so if $\sigma \in G$, then $\sigma(a) = \omega^i a$ for some i. We write $i \bmod n$ for the smallest nonnegative integer congruent to i modulo n. Let

$$\mathcal{S} = \left\{ i \bmod n : \sigma(a)/a = \omega^i \text{ for some } \sigma \in G \right\}.$$

Then $\mathcal{S}$ is the image of the function $G \to \mathbb{Z}/n\mathbb{Z}$ given by $\sigma \mapsto i \bmod n$, where $\sigma(a)/a = \omega^i$. This map is a well-defined group homomorphism whose image is $\mathcal{S}$, and it is injective, since if $\sigma \mapsto 0 \bmod n$, then $\sigma(a) = a$, so $\sigma = \mathrm{id}$. Therefore, $G \cong \mathcal{S}$, a subgroup of $\mathbb{Z}/n\mathbb{Z}$; hence, G is cyclic.

It remains to determine $|G|$ and $\min(F, a)$. Let $\mathrm{Gal}(K/F) = \langle \tau \rangle$, and set $\tau(a) = \omega^t a$. If $m = |G|$, then m is the least positive integer such that $(\omega^t)^m = 1$. The polynomial $\prod_{i=0}^{m-1}(x - \tau^i(a))$ lies in $F[x]$, since it is fixed by τ. Looking at the constant term, we see that $a^m \in F$. Therefore, $b^m = (a^m)^n \in F^{*n}$. If m' is the order of bF^{*n} in F^*/F^{*n}, then m' divides m. For the reverse divisibility, we know that $b^{m'} \in F^{*n}$, so $b^{m'} = c^n$ for some $c \in F$. Then $a^{m'n} = c^n$, so $a^{m'} = c\omega^i$ for some i, which means $a^{m'} \in F$. Therefore, $\tau^{m'}(a) = \omega^{tm'}a = a$, so m divides m', since m is the order of ω^t in F^*. Both divisibilities together yield $m = m'$. Moreover, since $m = [K : F] = \deg(\min(F, a))$ and $x^m - a^m \in F[x]$ has a as a root, we see that $\min(F, a) = x^m - a^m$. This finishes the proof. □

The simple structure of a cyclic group allows us to give a nice description of the intermediate fields of a cyclic extension. This description was hinted at in Example 5.4.

Corollary 9.7 *Let K/F be a cyclic extension of degree n, and suppose that F contains a primitive nth root of unity. If $K = F(\sqrt[n]{a})$ with $a \in F$, then any intermediate field of K/F is of the form $F(\sqrt[m]{a})$ for some divisor m of n.*

Proof. Let σ be a generator for $\mathrm{Gal}(K/F)$. Then any subgroup of $\mathrm{Gal}(K/F)$ is of the form $\langle \sigma^t \rangle$ for some divisor t of n. By the fundamental theorem, the intermediate fields are the fixed fields of the σ^t. If t is a divisor of n, write $n = tm$, and let $\alpha = \sqrt[n]{a}$. Then $\sigma^t(\alpha^m) = (\omega^t\alpha)^m = \alpha^m$, so α^m is fixed by σ^t. However, the order of $a^t F^{*n}$ in F^*/F^{*n} is m, so $F(\sqrt[m]{a})$ has degree m over F by Proposition 9.6. By the fundamental theorem, the fixed field of σ^t has degree m over F, which forces $F(\sqrt[m]{a})$ to be the fixed field of σ^t. This shows that any intermediate field of K/F is of the form $F(\sqrt[m]{a})$ for some divisor m of n. □

We now describe cyclic extensions of degree p in characteristic p. Let F be a field of characteristic $p > 0$. Define $\wp : F \longrightarrow F$ by $\wp(a) = a^p - a$. Then $\wp$ is an additive group homomorphism with kernel $\mathbb{F}_p$. To see this, if $a, b \in F$, then

$$\begin{aligned}\wp(a+b) &= (a+b)^p - (a+b) \\ &= a^p - a + b^p - b \\ &= \wp(a) + \wp(b),\end{aligned}$$

and $\wp(a) = 0$ if and only if $a^p = a$, if and only if $a \in \mathbb{F}_p$. Note that if $\wp(a) = b$, then $\wp(a+i) = b$ for all $i \in \mathbb{F}_p$, and in fact $\wp^{-1}(a) = \{a + i \mid i \in \mathbb{F}_p\}$. Therefore, if K is an extension of F such that there is an $\alpha \in K$ with $\wp(\alpha) = a \in F$, then $F(\alpha) = F(\wp^{-1}(a))$. The usual proof of the following theorem uses the additive version of Hilbert theorem 90, but, as with Lemma 9.4, we give a linear algebraic proof.

Theorem 9.8 *Let* $\operatorname{char}(F) = p$, *and let* K/F *be a cyclic Galois extension of degree* p. *Then* $K = F(\alpha)$ *with* $\alpha^p - \alpha - a = 0$ *for some* $a \in F$; *that is,* $K = F(\wp^{-1}(a))$.

Proof. Let σ be a generator of $\operatorname{Gal}(K/F)$, and let T be the linear transformation $T = \sigma - \mathrm{id}$. The kernel of T is

$$\begin{aligned}\ker(T) &= \{b \in K : \sigma(b) = b\} \\ &= F.\end{aligned}$$

Also, $T^p = (\sigma - \mathrm{id})^p = \sigma^p - \mathrm{id} = 0$, since the order of σ is p and $\operatorname{char}(F) = p$. Thus, $\operatorname{im}\left(T^{p-1}\right) \subseteq \ker(T)$. Because $\ker(T) = F$ and $\operatorname{im}\left(T^{p-1}\right)$ is an F-subspace of K, we get $\operatorname{im}\left(T^{p-1}\right) = \ker(T)$. Therefore, $1 = T^{p-1}(c)$ for some $c \in K$. Let $\alpha = T^{p-2}(c)$. Then $T(\alpha) = 1$, so $\sigma(\alpha) - \alpha = 1$ or $\sigma(\alpha) = \alpha + 1$. Since α is not fixed by σ, we see that $\alpha \notin F$, so $F(\alpha) = K$ because $[K : F] = p$ is prime. Now,

$$\begin{aligned}\sigma(\alpha^p - \alpha) &= \sigma(\alpha)^p - \sigma(\alpha) = (\alpha + 1)^p - (\alpha + 1) \\ &= \alpha^p + 1 - \alpha - 1 = \alpha^p - \alpha.\end{aligned}$$

If $a = \alpha^p - \alpha$, then $\wp(\alpha) = a \in F$, so $\alpha^p - \alpha - a = 0$. □

The converse of this theorem is also true.

Theorem 9.9 *Let* F *be a field of characteristic* p, *and let* $a \in F - \wp^{-1}(F)$. *Then* $f(x) = x^p - x - a$ *is irreducible over* F, *and the splitting field of* f *over* F *is a cyclic Galois extension of* F *of degree* p.

Proof. Let K be the splitting field of f over F. If α is a root of f, it is easy to check that $\alpha + 1$ is also a root of f. Hence, the p roots of f are $\alpha, \alpha + 1, \ldots, \alpha + p - 1$. Therefore, $K = F(\alpha)$. The assumption on a assures us that $\alpha \notin F$. Assume for now that f is irreducible over F. Then $[K : F] = \deg(f) = p$. By the isomorphism extension theorem, there is a $\sigma \in \operatorname{Gal}(K/F)$ with $\sigma(\alpha) = \alpha + 1$. From this, it follows that the order of σ is p, so $\operatorname{Gal}(K/F) = \langle\sigma\rangle$. This proves that K/F is a cyclic Galois extension.

It remains for us to prove that $f(x)$ is irreducible over F. If not, then f factors over F as $f(x) = g_1(x) \cdots g_r(x)$, with each g_i irreducible over F. If β is a root of g_i for some i, then the paragraph above shows that $K = F(\beta)$, so $[K : F] = \deg(g_i)$. This forces all degrees of the g_i to be the same, so $\deg(f) = r \deg(g_1)$. Since $\deg(f)$ is prime and f does not split over F, we see that $r = 1$; hence, f is irreducible over F. □

Example 9.10 Let $F = \mathbb{F}_p(x)$ be the rational function field in one variable over $\mathbb{F}_p$. We claim that $x \notin \wp^{-1}(F)$, so the extension $F(\wp^{-1}(x))$ is a cyclic extension of F of degree p. To prove this, suppose instead that $x \in \wp^{-1}(F)$, so $x = a^p - a$ for some $a \in F$. We can write $a = f/g$ with $f, g \in \mathbb{F}_p[x]$

relatively prime. Then $x = f^p/g^p - f/g$, or $g^p x = f^p - fg^{p-1}$. Solving for f^p gives $f^p = g^{p-1}(gx - f)$, so g divides f^p. This is impossible; thus, $x \notin \wp^{-1}(F)$, and then $F(\wp^{-1}(F))$ is a cyclic extension of F of degree p as we claimed.

Problems

1. Suppose that F is a field containing a primitive nth root of unity, and let $a \in F$. Show that $x^n - a$ is irreducible over F if and only if a is not an mth power for any $m > 1$ dividing n.

2. Suppose that F is a field, and let ω be a primitive nth root of unity in an algebraic closure of F. If $a \in F$ is not an mth power in $F(\omega)$ for any $m > 1$ that divides n, show that $x^n - a$ is irreducible over F.

3. This problem describes cyclic extensions of degree four of a base field that does not contain a primitive fourth root of unity. Let F be a field that does not contain a primitive fourth root of unity. Let $L = F(\sqrt{a})$ for some $a \in F - F^2$, and let $K = L(\sqrt{b})$ for some $b \in L - L^2$. Show that the following statements are equivalent:

   (a) a is a sum of two squares in F.

   (b) $-1 = N_{L/F}(\alpha)$ for some $\alpha \in L$.

   (c) $a = N_{L/F}(\alpha)$ for some $\alpha \in L$.

   (d) $N_{L/F}(b) \equiv a \bmod F^{*2}$ for some $b \in L$.

   (e) K/F is a cyclic extension (with the b in Problem 3d).

   (f) L lies in a cyclic extension of F of degree 4.

4. This problem investigates the splitting field of the polynomial $x^n - a$ over a field F that does not contain a primitive nth root of unity.

   (a) If $a \in F$, show that the splitting field of $x^n - a$ over F is $F(\alpha, \omega)$, where $\alpha^n = a$ and ω is a primitive nth root of unity.

   (b) Let $N = F(\alpha, \omega)$, let $K = F(\alpha)$, and let $L = F(\omega)$. Show that L/F is Galois and N/L is cyclic.

   (c) Suppose that $\min(F, \omega) = (x - \omega)(x - \omega^{-1})$ and that $[N : L] = n$. Show that there is an element $\sigma \in \text{Gal}(N/F)$ with $\sigma(\alpha) = \omega\alpha$ and $\sigma(\omega) = \omega$, and a τ with $\tau(\omega) = \omega^{-1}$ and $\tau(\alpha) = \alpha$. Moreover, show that the order of σ is n, the order of τ is 2, and $\tau\sigma\tau = \sigma^{-1}$. Recall the definition of the *dihedral group* D_n, and show that $D_n = \text{Gal}(N/F)$.

(d) Let p be an odd prime, and let $\omega \in \mathbb{C}$ be a primitive pth root of unity. Let $F = \mathbb{Q}(\omega) \cap \mathbb{R}$. Let $a \in \mathbb{Q}$ be a rational number that is not a pth power in $\mathbb{Q}$. Show that $[F(\sqrt[p]{a}) : F] = p$ and that if $L = F(\omega)$, then $[L(\sqrt[p]{a}) : L] = p$. Conclude that if N is the splitting field of $x^p - a$ over F, then $\mathrm{Gal}(N/F) = D_p$.

5. In this problem, we prove the following result: Suppose that K/F is a finite extension with K algebraically closed. Then $\mathrm{char}(F) = 0$ and $K = F(\sqrt{-1})$. Use the following steps to prove this:

(a) If $\mathrm{char}(F) = p > 0$ and $\beta \in F - F^p$, then $x^{p^r} - \beta$ is irreducible over F for all $r > 0$.

(b) If $\mathrm{char}(F) = p > 0$ and there is a cyclic extension of degree p, then there are cyclic extensions of F of degree p^r for any $r \geq 1$.

(c) Let p be a prime, and suppose that either F contains a primitive pth root of unity for p odd, or that F contains a primitive fourth root of unity for $p = 2$. If there is an $a \in F$ with $x^p - a$ irreducible over F, then $x^{p^2} - a$ is irreducible over F.
(Hint: Use a norm argument.)

(d) Use the previous steps to prove the result.

10 Hilbert Theorem 90 and Group Cohomology

In this section, we change gears. Instead of investigating Galois extensions with certain types of Galois groups, we investigate some deep ideas that arise in classical treatments of cyclic Galois extensions. Cohomology, first introduced in algebraic topology, is a valuable tool in many areas of algebra, including group theory, the theory of algebras, and algebraic geometry. We introduce the notions of group cohomology here, we give a couple of applications of the theory, and we relate it to cyclic extensions. To start with, we prove the so-called Hilbert theorem 90, which can be used to prove Lemma 9.4, the key step in characterizing cyclic extensions.

In order to prove the Hilbert theorem 90, we define a concept that we will see again when we formally define group cohomology. Let K be a field, and let G be a subgroup of $\mathrm{Aut}(K)$. A *crossed homomorphism* $f : G \longrightarrow K^*$ is a function that satisfies $f(\sigma\tau) = f(\sigma) \cdot \sigma(f(\tau))$ for all $\sigma, \tau \in G$.

Proposition 10.1 *Let K be a Galois extension of F with Galois group G, and let $f : G \longrightarrow K^*$ be a crossed homomorphism. Then there is an $a \in K$ with $f(\tau) = \tau(a)/a$ for all $\sigma \in G$.*

Proof. The Dedekind independence lemma shows that $\sum_{\sigma \in G} f(\sigma)\sigma(c) \neq 0$ for some $c \in K$, since each $f(\sigma) \neq 0$. Let $b = \sum_{\sigma \in G} f(\sigma)\sigma(c)$. Then

$\tau(b) = \sum_{\sigma \in G} \tau(f(\sigma))(\tau\sigma)(c)$, so

$$\begin{aligned} f(\tau)\tau(b) &= \sum_{\sigma \in G} f(\tau)\tau(f(\sigma)) \cdot (\tau\sigma)(c) \\ &= \sum_{\sigma \in G} f(\tau\sigma) \cdot (\tau\sigma)(c) = b. \end{aligned}$$

Thus, $f(\tau) = b/\tau(b)$. Setting $a = b^{-1}$ proves the result. □

Theorem 10.2 (Hilbert Theorem 90) *Let K/F be a cyclic Galois extension, and let σ be a generator of* Gal(K/F). *If $u \in K$, then $N_{K/F}(u) = 1$ if and only if $u = \sigma(a)/a$ for some $a \in K$.*

Proof. One direction is easy. If $u = \sigma(a)/a$, then $N_{K/F}(\sigma(a)) = N_{K/F}(a)$, so $N(u) = 1$. Conversely, if $N_{K/F}(u) = 1$, then define $f : G \longrightarrow K^*$ by $f(\text{id}) = 1$, $f(\sigma) = u$, and $f(\sigma^i) = u\sigma(u)\cdots\sigma^{i-1}(u)$ for $i < n$. To show that f is a crossed homomorphism, let $0 \le i, j < n$. If $i + j < n$, then

$$\begin{aligned} f(\sigma^i\sigma^j) &= f(\sigma^{i+j}) = u\sigma(u)\cdots\sigma^{i+j-1}(u) \\ &= \left(u\sigma(u)\cdots\sigma^{i-1}(u)\right) \cdot \sigma^i \left(u\sigma(u)\cdots\sigma^{i-1}(u)\right) \\ &= f(\sigma^i) \cdot \sigma^i(f(\sigma^j)). \end{aligned}$$

If $i + j \ge n$, then $0 \le i + j - n < n$, so

$$f(\sigma^i\sigma^j) = f(\sigma^{i+j}) = f(\sigma^{i+j-n}) = u\sigma(u)\cdots\sigma^{i+j-n-1}(u).$$

However,

$$\begin{aligned} f(\sigma^i)\sigma^i(f(\sigma^j)) &= \left(u\sigma(u)\cdots\sigma^{i-1}(u)\right) \cdot \sigma^i \left(u\sigma(u)\cdots\sigma^{j-1}(u)\right) \\ &= \left(u\sigma(u)\cdots\sigma^{i+j-n-1}(u)\right) \cdot \sigma^{i+j-n} \left(u\sigma(u)\cdots\sigma^{n-1}(u)\right) \\ &= f(\sigma^i\sigma^j) \cdot N_{K/F}(u) \\ &= f(\sigma^i\sigma^j). \end{aligned}$$

Therefore, f is a crossed homomorphism. By Proposition 10.1, there is an $a \in K$ with $f(\sigma^i) = \sigma^i(a)/a$ for all i. Thus, $u = f(\sigma) = \sigma(a)/a$. □

Lemma 9.4 follows quickly from the Hilbert theorem 90. If K/F is a cyclic extension of degree n, if σ is a generator of Gal(K/F), and if F contains a primitive nth root of unity ω, then $N_{K/F}(\omega) = \omega^n = 1$. Therefore, $\omega = \sigma(a)/a$ for some $a \in K$. This gives an alternative proof of Lemma 9.4, the proof most commonly seen in Galois theory texts.

Group cohomology

We can mimic the arguments above to get results about the trace. However, before we do so, we introduce group cohomology. Given a group G and an Abelian group M with some extra structure to be described shortly, we will obtain a sequence of *cohomology groups* $H^n(G, M)$, one for each nonnegative integer.

Let G be a group, and let M be an Abelian group. We say that M is a *G-module* if there is a function $G \times M \to M$, where the image of (σ, m) is written σm, such that

$$\begin{aligned} 1m &= m, \\ \sigma(\tau m) &= (\sigma\tau)m, \\ \sigma(m_1 + m_2) &= \sigma m_1 + \sigma m_2 \end{aligned}$$

for all $m, m_1, m_2 \in M$ and all $\sigma, \tau \in G$. This is equivalent to the condition that M is a left module over the *group ring* $\mathbb{Z}[G]$. For example, if K is a Galois extension of a field F and $G = \mathrm{Gal}(K/F)$, then K^* is a G-module by defining $\sigma a = \sigma(a)$. Similarly, the additive group $(K, +)$ is a G-module.

Suppose that M is a G-module. Let $C^n(G, M)$ be the set of all functions from the Cartesian product $G \times G \times \cdots \times G$ (n times) to M. The elements of $C^n(G, M)$ are called *n-cochains*. If $n = 0$, we define $C^0(G, M) = M$. The set $C^n(G, M)$ can be made into a group by adding functions componentwise; that is, if $f, g \in C^n(G, M)$, define $f + g$ by

$$(f + g)(\sigma_1, \dots, \sigma_n) = f(\sigma_1, \dots, \sigma_n) + g(\sigma_1, \dots, \sigma_n).$$

One can easily check that with this operation $C^n(G, M)$ is an Abelian group. Note that $C^n(G, M) = \hom_{\mathbb{Z}}(\mathbb{Z}[G^n], M)$, which is another way to see that $C^n(G, M)$ is an Abelian group.

Define a map $\delta_n : C^n(G, M) \to C^{n+1}(G, M)$ by

$$\begin{aligned} \delta_n(f)(\sigma_1, \dots, \sigma_{n+1}) = \sigma_1 f(\sigma_2, \dots, \sigma_{n+1}) & \\ + \sum_{i=1}^{n} (-1)^i f(\sigma_1, \dots, \sigma_i\sigma_{i+1}, \dots, \sigma_{n+1}) & \\ + (-1)^{n+1} f(\sigma_1, \dots, \sigma_n). & \end{aligned}$$

If $n = 0$, then the map $\delta_0 : M = C^0(G, M) \to C^1(G, M)$ is defined by $\delta_0(m)(\sigma) = \sigma m - m$. This definition is compatible with the general formula above. A straightforward but tedious calculation shows that δ_n is a homomorphism and that $\delta_{n+1} \circ \delta_n$ is the zero map (see Problems 1 and 2). The maps δ_n are called *boundary maps*.

Let $Z^n(G, M) = \ker(\delta_n)$. The elements of $Z^n(G, M)$ are called *n-cocycles*. Since $\delta_n(\delta_{n-1}(f)) = 0$ for all $f \in C^{n-1}(G, M)$, the image of δ_{n-1}

is contained in $\ker(\delta_n)$. Let $B^n(G, M) = \mathrm{im}(\delta_{n-1})$ if $n > 0$. For $n = 0$, let $B^0(G, M) = 0$. The elements of $B^n(G, M)$ are called *n-coboundaries*. Finally, the *nth cohomology group* $H^n(G, M)$ of G with coefficients in M is defined by

$$H^n(G, M) = Z^n(G, M)/B^n(G, M).$$

Two cocycles in $Z^n(G, M)$ are said to be *cohomologous* if they represent the same element in $H^n(G, M)$; that is, if they differ by a coboundary.

Let us look at the cohomology groups for small n. The kernel of δ_0 consists of all $m \in M$ with $\sigma m = m$ for all $\sigma \in G$. Therefore,

$$H^0(G, M) = M^G = \{m \in M : \sigma m = m \text{ for all } \sigma \in G\}.$$

If $n = 1$, then $f : G \longrightarrow M$ is a 1-cocycle if $\delta_1(f) = 0$. This happens when $\sigma f(\tau) - f(\sigma\tau) + f(\sigma) = 0$ for all $\sigma, \tau \in G$. In other words, a 1-cocycle is a crossed homomorphism as defined above, at least when M is the multiplicative group of a field. If g is a 1-coboundary, then there is an $m \in M$ with $g(\sigma) = \sigma m - m$ for all $\sigma \in G$. Proposition 10.1 implies that if $G = \mathrm{Gal}(K/F)$, then any 1-cocycle from G to K^* is a 1-coboundary. In other words, $H^1(G, K^*) = 0$. This result is often referred to as the cohomological Hilbert theorem 90. It is also true that $H^1(G, K) = 0$, as we now prove.

Proposition 10.3 *Let K/F be a Galois extension with Galois group G, and let $g : G \longrightarrow K$ be a 1-cocycle. Then there is an $a \in K$ with $g(\tau) = \tau(a) - a$ for all $\tau \in G$.*

Proof. Since K/F is separable, the trace map $T_{K/F}$ is not the zero map. Thus, there is a $c \in K$ with $T_{K/F}(c) \neq 0$. If $\alpha = T_{K/F}(c)$, then $\alpha \in F^*$ and $T_{K/F}(\alpha^{-1}c) = 1$. By replacing c with $\alpha^{-1}c$, we may assume that $T_{K/F}(c) = 1$. Recall that $T_{K/F}(x) = \sum_{\sigma \in G} \sigma(x)$ for all $x \in K$. Let $b = \sum_{\sigma \in G} g(\sigma)\sigma(c)$. Then $\tau(b) = \sum_{\sigma \in G} \tau(g(\sigma))(\tau\sigma)(c)$. Since $g(\tau\sigma) = g(\tau) + \tau(g(\sigma))$,

$$\begin{aligned}
\tau(b) &= \sum_{\sigma \in G} (g(\tau\sigma) - g(\tau))\,(\tau\sigma)(c) \\
&= \sum_{\sigma \in G} g(\tau\sigma)(\tau\sigma)(c) - \sum_{\sigma \in G} g(\tau)(\tau\sigma)(c) \\
&= b - g(\tau) \cdot \tau\left(\sum_{\sigma \in G} \sigma(c)\right) \\
&= b - g(\tau).
\end{aligned}$$

Therefore, $g(\tau) = b - \tau(b)$. Setting $a = -b$ gives $g(\tau) = \tau(a) - a$ for all $\tau \in G$. □

We record our two results about H^1 in the following corollary.

Corollary 10.4 (Cohomological Hilbert Theorem 90) *Let K be a Galois extension of F with Galois group G. Then $H^1(G, K^*) = 0$ and $H^1(G, K) = 0$.*

The triviality of $H^1(G, K)$ can be used to give information about the trace map of a cyclic extension and to give an alternative proof of Theorem 9.8, the proof that is typically seen in texts. We now obtain the analog of the Hilbert theorem 90 for the trace map.

Theorem 10.5 (Additive Hilbert Theorem 90) *Let K be a cyclic Galois extension of F, and let σ be a generator of* $\mathrm{Gal}(K/F)$. *If $u \in K$, then $T_{K/F}(u) = 0$ if and only if $u = \sigma(a) - a$ for some $a \in K$.*

Proof. If $u = \sigma(a) - a$, then $T_{K/F}(u) = 0$. Conversely, suppose that $T_{K/F}(u) = 0$. Let $n = [K : F]$, and define $g : G \longrightarrow K$ by $g(\mathrm{id}) = 0$, $g(\sigma) = u$, and for $i < n$ by

$$g(\sigma^i) = u + \sigma(u) + \cdots + \sigma^{i-1}(u).$$

If $0 \le i, j < n$, then as $0 = T_{K/F}(u) = \sum_{i=1}^n \sigma^i(u)$, we see that regardless of whether $i + j < n$ or $i + j \ge n$, we have

$$\begin{aligned} g(\sigma^i \sigma^j) &= u + \sigma(u) + \cdots + \sigma^{i+j-1}(u) \\ &= \left(u + \sigma(u) + \cdots + \sigma^{i-1}(u)\right) + \sigma^i\left(u + \sigma(u) + \cdots + \sigma^{j-1}(u)\right) \\ &= g(\sigma^i) + \sigma^i(g(\sigma^j)). \end{aligned}$$

Therefore, g is a cocycle. By Proposition 10.3, there is an $a \in K$ with $g(\sigma^i) = \sigma^i(a) - a$ for all i. Hence, $u = g(\sigma) = \sigma(a) - a$. □

The usual argument for Theorem 9.8 goes as follows. If K/F is a cyclic extension of degree p with $\mathrm{char}(F) = p$, then $T_{K/F}(1) = p \cdot 1 = 0$, so by the additive Hilbert theorem 90, $1 = \sigma(a) - a$ for some $a \in K$. It is then easy to see that a is a root of $x^p - x - c$ for some $c \in F$ and that $K = F(a)$.

Group extensions

Second cohomology groups have some important applications. In what follows, we will discuss applications to group theory and to the theory of division algebras. Before doing so, we write out the formulas that determine when a 2-cochain is a 2-cocycle or a 2-coboundary. Let G be a group, and let M be a G-module. A function $f : G \times G \to M$ is a 2-cocycle if for each $\sigma, \tau, \rho \in G$, we have

$$f(\sigma, \tau) f(\sigma\tau, \rho) = \sigma f(\tau, \rho) f(\sigma, \tau\rho).$$

We will refer to this equation as the *cocycle condition.* On the other hand, if there are $m_\sigma \in M$ with

$$f(\sigma, \tau) = m_\sigma + \sigma m_\tau - m_{\sigma\tau}$$

for each $\sigma, \tau \in G$, then f is a 2-coboundary.

The first application of second cohomology groups we give is to group extensions. We point out that a number of statements in the remainder of this section will be left as exercises. Suppose that E is a group that contains an Abelian normal subgroup M, and let $G = E/M$. We then say that E is a *group extension* of G by M. The basic problem is this: Given groups G and M, describe all groups E that, up to isomorphism, contain M as a normal subgroup and have $E/M \cong G$. As we shall see, if M is Abelian, then $H^2(G, M)$ classifies group extensions of G by M.

Example 10.6 Let $E = S_3$. If $M = \langle (123) \rangle$, then M is isomorphic to $\mathbb{Z}/3\mathbb{Z}$ and M is an Abelian normal subgroup of E. The quotient group E/M is isomorphic to $\mathbb{Z}/2\mathbb{Z}$. Therefore, S_3 is a group extension of $\mathbb{Z}/2\mathbb{Z}$ by $\mathbb{Z}/3\mathbb{Z}$.

Example 10.7 Let $E = D_n$, the dihedral group. One description of E is by generators and relations. The group E is generated by elements σ and τ satisfying $\tau^n = \sigma^2 = e$ and $\sigma\tau\sigma = \tau^{-1}$. Let $M = \langle \sigma \rangle$, a normal subgroup of E that is isomorphic to $\mathbb{Z}/n\mathbb{Z}$. The quotient E/M is isomorphic to $\mathbb{Z}/2\mathbb{Z}$, so E is a group extension of $\mathbb{Z}/2\mathbb{Z}$ by $\mathbb{Z}/n\mathbb{Z}$.

Example 10.8 Let M and G be groups, and let $\varphi : G \to \mathrm{End}(M)$ be a group homomorphism. If E is the *semidirect product* $M \times_\varphi G$, then $M' = \{(m, e) : m \in M\}$ is a normal subgroup of E isomorphic to M, and $E/M' \cong G$. Thus, E is a group extension of M by G. Notice that the group extensions in each of the two previous examples are also semidirect products.

Suppose that M is Abelian and that E is a group extension of G by M. We can make M into a G-module as follows. View $G = E/M$. If $\sigma \in G$ and $m \in M$, let e be any element of E that is a coset representative of σ. Then define $\sigma m = eme^{-1}$. Note that we will write the group operations in these groups multiplicatively. The groups G and E need not be Abelian, although we are assuming that M is Abelian. It is not hard to show that this definition gives a well-defined action of G on M and that M is a G-module. We can obtain a 2-cocycle from this information. For each $\sigma \in G$, pick a coset representative $e_\sigma \in E$. The map $\sigma \mapsto e_\sigma$ need not be a homomorphism. Let $f(\sigma, \tau) = e_\sigma e_\tau e_{\sigma\tau}^{-1}$. Then the coset of $f(\sigma, \tau)$ in G is trivial, so $e_\sigma e_\tau e_{\sigma\tau}^{-1} \in M$. Therefore, f is a function from $G \times G$ to M. Moreover, a short calculation shows that f is actually a 2-cocycle. The cocycle f does depend on the choice of coset representatives chosen. Suppose that $\{d_\sigma\}$ is another set of coset representatives for the elements of G. Then there are $m_\sigma \in M$ with $d_\sigma = m_\sigma e_\sigma$. Let g be the cocycle obtained by the choice of the d_σ; that is, $g(\sigma, \tau) = d_\sigma d_\tau d_{\sigma\tau}^{-1}$. Then

$$g(\sigma, \tau) = d_\sigma d_\tau d_{\sigma\tau}^{-1} = (m_\sigma e_\sigma)(m_\tau e_\tau)(m_{\sigma\tau} e_{\sigma\tau})^{-1}$$

$$
\begin{aligned}
&= m_\sigma \sigma m_\tau e_\sigma e_\tau e_{\sigma\tau}^{-1} m_{\sigma\tau}^{-1} \\
&= (m_\sigma \sigma m_\tau m_{\sigma\tau}^{-1}) e_\sigma e_\tau e_{\sigma\tau}^{-1} \\
&= (m_\sigma \sigma m_\tau m_{\sigma\tau}^{-1}) f(\sigma, \tau).
\end{aligned}
$$

In this calculation, we used the fact that $e_\sigma m e_\sigma^{-1} = \sigma m$. The function $(\sigma, \tau) \mapsto m_\sigma \sigma m_\tau m_{\sigma\tau}^{-1}$ is the image under δ_1 of the 1-cochain $\sigma \mapsto m_\sigma$. Therefore, f and g differ by a 1-coboundary, so they determine the same element of $H^2(G, M)$. We have thus shown that for any group extension E of G and M there is a uniquely determined element of $H^2(G, M)$.

We can reverse these calculations. Let M be a G-module and let $f \in Z^2(G, M)$. We can define a group E_f as follows. As a set, $E_f = M \times G$. However, multiplication in E_f is defined by

$$(m, \sigma)(n, \tau) = (m \cdot \sigma n \cdot f(\sigma, \tau), \sigma\tau).$$

A short calculation shows that this is an associative operation with an identity $(f(1,1)^{-1}, 1)$, and $(m, \sigma)^{-1} = (m^{-1} f(1,1)^{-1}, \sigma^{-1})$. In fact, associativity follows exactly from the condition that f is a 2-cocycle. The formulas for identity and inverses use the fact that $f(1,1) = f(1, \sigma) = f(\sigma, 1)$ for any $\sigma \in G$, which also follows from the cocycle condition. The group M is isomorphic to the normal subgroup $\{(m, 1) : m \in M\}$ of E_f, and the quotient of E_f by this subgroup is isomorphic to G. It is not hard to show that if g is another 2-cocycle that differs from f by a 2-coboundary, then the resulting group obtained from g is isomorphic to E_f. By being more precise about the definition of a group extension, these arguments would then show that the group extensions of M by G are classified by $H^2(G, M)$.

Example 10.9 Let M and G be groups and $\varphi : G \to \mathrm{End}(M)$ be a group homomorphism. Let $E = M \times_\varphi G$ be the semidirect product of M by G. We determine the cocycle describing E. Let $M' = \{(m, e) : m \in M\}$ and $G' = \{(e, g) : g \in G\}$ be the isomorphic copies of M and G inside E. The elements of G' form a natural set of coset representatives of M' in E. The cocycle f describing E is defined by

$$f(\sigma, \tau) = (e, \sigma)(e, \tau)(e, \sigma\tau)^{-1} = (e, e),$$

so f is the trivial cocycle.

Conversely, if f is the trivial cocycle of $H^2(G, M)$, then we can see that the group extension constructed from G and M and f is a semidirect product of M by G, for the mapping $\sigma \mapsto e_\sigma$ defined earlier is a homomorphism if and only if the corresponding cocycle is trivial. Since this map is a homomorphism, we can check that the map $\varphi : G \to \mathrm{End}(M)$, where $\varphi(\sigma)$ is the automorphism $m \mapsto e_\sigma m e_\sigma^{-1}$, is also a homomorphism, and the group E_f constructed above from G, M, and f is the semidirect product $M \times_\varphi G$.

Example 10.10 Let Q_8 be the quaternion group. Then $Q_8 = \{\pm 1, \pm i, \pm j, \pm k\}$, and the operation on Q_8 is given by the relations $i^2 = j^2 = k^2 = -1$ and $ij = k = -ji$. We show that Q_8 is a group extension of $M = \langle i \rangle$ by $\mathbb{Z}/2\mathbb{Z}$, and we determine the cocycle for this extension. First note that M is an Abelian normal subgroup of Q_8 and that $Q_8/M \cong \mathbb{Z}/2\mathbb{Z}$. Therefore, Q_8 is a group extension of M by $\mathbb{Z}/2\mathbb{Z}$. We use 1 and j as coset representatives of M in Q_8. Our cocycle f that represents this group extension is then given by

$$\begin{aligned} f(1,1) &= f(1,j) = f(j,1) = 1, \\ f(j,j) &= j^2 = -1. \end{aligned}$$

This cocycle is not trivial, so Q_8 is not the semidirect product of M and $\mathbb{Z}/2\mathbb{Z}$. In fact, Q_8 is not the semidirect product of any two subgroups, because one can show that there do not exist two subgroups of Q_8 whose intersection is $\langle 1 \rangle$.

Crossed products

Another application of the second cohomology group is in the theory of algebras. If F is a field, then an *F-algebra* is a ring A that is also an F-vector space, in which multiplication in A and scalar multiplication are connected by the axiom

$$\alpha(ab) = (\alpha a)b = a(\alpha b)$$

for all $a, b \in A$ and all $\alpha \in F$. Let K be a Galois extension of F with Galois group G. If $f \in Z^2(G, K^*)$, we can construct an F-algebra from K, G, and f as follows. For each $\sigma \in G$, let x_σ be a symbol and let A be the Abelian group

$$A = \oplus_{\sigma \in G} K x_\sigma.$$

We can define multiplication on A by using the two definitions

$$\begin{aligned} x_\sigma x_\tau &= f(\sigma, \tau) x_{\sigma\tau}, \\ x_\sigma a &= \sigma(a) x_\sigma. \end{aligned}$$

A full definition of multiplication can then be obtained by using distributivity; that is,

$$\sum_{\sigma \in G} a_\sigma x_\sigma \cdot \sum_{\tau \in G} b_\tau x_\tau = \sum_{\sigma, \tau \in G} a_\sigma \sigma(b_\tau) f(\sigma, \tau) x_{\sigma\tau}.$$

A calculation shows that associativity of multiplication follows immediately from the cocycle condition and that the other axioms of an F-algebra are straightforward. The algebra A is an F-vector space of dimension $|G| \cdot [K : F] = |G|^2$. This algebra is called a *crossed product* and is often denoted $A =$

$(K/F, G, f)$. Crossed products come up in the theory of division algebras. It is known that any crossed product is isomorphic to a ring of $n \times n$ matrices over a division ring. Moreover, if D is a division ring that is finite dimensional over the field $F = \{a \in D : da = ad \text{ for all } d \in D\}$, the center of D, then some matrix ring over D is isomorphic to a crossed product algebra of the form $(K/F, G, f)$ for some Galois extension K of F.

The algebra A is determined up to isomorphism not by the cocycle f but by the class of f in $H^2(G, M)$, as we now show. Suppose that g is another 2-cocycle that differs from f by a 2-coboundary. Then there are $a_\sigma \in K^*$ with $g(\sigma, \tau) = a_\sigma \sigma(a_\tau) a_{\sigma\tau}^{-1} f(\sigma, \tau)$. Let $y_\sigma = a_\sigma x_\sigma$. Then $K y_\sigma = K x_\sigma$, so $A = \oplus_{\sigma \in G} K y_\sigma$. Moreover, $y_\sigma a = \sigma(a) y_\sigma$ for all $a \in K$, and

$$\begin{aligned} y_\sigma y_\tau y_{\sigma\tau}^{-1} &= a_\sigma x_\sigma a_\tau x_\tau (a_{\sigma\tau} x_{\sigma\tau})^{-1} \\ &= a_\sigma \sigma(a_\tau) x_\sigma x_\tau x_{\sigma\tau}^{-1} a_{\sigma\tau}^{-1} \\ &= a_\sigma \sigma(a_\tau) a_{\sigma\tau}^{-1} f(\sigma, \tau) \\ &= g(\sigma, \tau). \end{aligned}$$

Therefore, the algebra constructed with the procedure above using the cocycle g is isomorphic to A. Conversely, if the algebras constructed from two cocycles are isomorphic, then it can be seen that the cocycles are cohomologous; that is, they represent the same element in $H^2(G, M)$.

Example 10.11 Let $\mathbb{H}$ be Hamilton's *quaternions*. The ring $\mathbb{H}$ consists of all symbols $a + bi + cj + dk$ with $a, b, c, d \in \mathbb{R}$, and multiplication is given by the relations $i^2 = j^2 = k^2 = -1$ and $ij = k = -ji$. This was the first example of a noncommutative division ring. The field of complex numbers $\mathbb{C}$ can be viewed as the subring of $\mathbb{H}$ consisting of all elements of the form $a + bi$, and $\mathbb{H} = \mathbb{C} \oplus \mathbb{C}j$. The extension $\mathbb{C}/\mathbb{R}$ is Galois with Galois group $\{\mathrm{id}, \sigma\}$, where σ is complex conjugation. Let $x_{\mathrm{id}} = 1$ and $x_\sigma = j$. Then

$$x_\sigma (a + bi) x_\sigma^{-1} = j(a + bi) j^{-1} = a - bi = \sigma(a + bi).$$

The cocycle f associated to this algebra is given by

$$\begin{aligned} f(\mathrm{id}, \mathrm{id}) &= x_{\mathrm{id}} x_{\mathrm{id}} x_{\mathrm{id}}^{-1} = 1, \\ f(\mathrm{id}, \sigma) &= x_{\mathrm{id}} x_\sigma x_\sigma^{-1} = 1, \\ f(\sigma, \mathrm{id}) &= x_\sigma x_{\mathrm{id}} x_\sigma^{-1} = 1, \\ f(\sigma, \sigma) &= x_\sigma x_\sigma x_{\mathrm{id}}^{-1} = j^2 = -1. \end{aligned}$$

On the other hand, if we start with this cocycle and construct the crossed product $A = (\mathbb{C}/\mathbb{R}, \mathrm{Gal}(\mathbb{C}/\mathbb{R}), f)$, then $A = \mathbb{C}x_{\mathrm{id}} \oplus \mathbb{C}x_\sigma$, and the map $A \to \mathbb{H}$ given by $cx_{\mathrm{id}} + dx_\sigma \mapsto c + dj$ is an isomorphism of $\mathbb{R}$-algebras.

Example 10.12 Let K/F be a Galois extension of degree n with Galois group G, and consider the crossed product $A = (K/F, G, 1)$, where 1 represents the trivial cocycle. We will show that $A \cong M_n(F)$, the ring of $n \times n$

matrices over F. First, note that $A = \oplus_{\sigma\in G} K x_\sigma$, where multiplication on A is determined by the relations $x_\sigma x_\tau = x_{\sigma\tau}$ and $x_\sigma a = \sigma(a)x_\sigma$ for $a \in K$. If $f = \sum a_\sigma x_\sigma \in A$, then f induces a map $\varphi_f : K \to K$ given by $\varphi_f(k) = \sum a_\sigma \sigma(k)$. In other words, φ_f is the linear combination $\sum a_\sigma \sigma$. Each σ is an F-linear transformation of K, so $\varphi_f \in \mathrm{End}_F(K)$. The relations governing multiplication in A show that the map $\varphi : A \to \mathrm{End}_F(K)$ given by $\varphi(f) = \varphi_f$ is an F-algebra homomorphism. Moreover, φ is injective since if $\sum a_\sigma \sigma$ is the zero transformation, then each $a_\sigma = 0$ by the Dedekind independence lemma. Both A and $\mathrm{End}_F(K)$ have dimension n^2 over F, so φ is automatically surjective. This proves that $A \cong \mathrm{End}_F(K)$, and so $A \cong M_n(F)$.

Crossed products have a simpler description when we start with a cyclic extension. In addition, the norm map helps to describe crossed products in this situation. Suppose that K/F is a cyclic Galois extension with Galois group $G = \langle\sigma\rangle$, and let $a \in F^*$. We can define a cocycle in $H^2(G, K^*)$ by

$$f(\sigma^i, \sigma^j) = \begin{cases} 1 & \text{if } i+j < n \\ a & \text{if } i+j \geq n. \end{cases}$$

A straightforward calculation shows that f is indeed a 2-cocycle. The algebra constructed from f is usually denoted $(K/F, \sigma, a)$ and is called a *cyclic algebra*. This construction is a special case of the crossed product construction. If $x = x_\sigma$, then $x\alpha x^{-1} = \sigma(\alpha)$ for all $\alpha \in K$, and $x^n = a$. These relations along with K and σ fully determine the algebra $(K/F, \sigma, a)$. If $a = N_{K/F}(c)$ for some $c \in K$, then if we set $y_\sigma = c^{-1}x_\sigma$, a short calculation shows that $y_\sigma^n = 1$. Therefore, the cocycle associated to y_σ is trivial, so $(K/F, \sigma, a) \cong M_n(F)$ by Example 10.12. Moreover, Problem 16 proves that $H^2(G, K^*) \cong F^*/N_{K/F}(K^*)$. One consequence of this fact is that two algebras $(K/F, \sigma, a)$ and $(K/F, \sigma, b)$ are isomorphic if and only if $ab^{-1} \in N_{K/F}(K^*)$. Moreover, by a theorem of the theory of algebras, if none of the elements $a, a^2, \ldots, a^{n-1}$ are equal to the norm from K to F of a nonzero element of K, then $(K/F, \sigma, a)$ is a division algebra. Hamilton's quaternions are of the form $(\mathbb{C}/\mathbb{R}, \sigma, -1)$.

The interested reader can find much more information about group extensions and crossed products in Rotman [23] and Jacobson [16].

Problems

1. Let M be a G-module. Show that the boundary map $\delta_n : C^n(G, M) \to C^{n+1}(G, M)$ defined in this section is a homomorphism.

2. With notation as in the previous problem, show that $\delta_{n+1} \circ \delta_n$ is the zero map.

3. Let M be a G-module, and let $f \in Z^2(G, M)$. Show that $f(1, 1) = f(1, \sigma) = f(\sigma, 1)$ for all $\sigma \in G$.

4. If E is a group with an Abelian normal subgroup M, and if $G = E/M$, show that the action of G on M given by $\sigma m = eme^{-1}$ if $eM = \sigma$ is well defined and makes M into a G-module.

5. With E, M, G as in the previous problem, if e_σ is a coset representative of σ, show that the function f defined by $f(\sigma, \tau) = e_\sigma e_\tau e_{\sigma\tau}^{-1}$ is a 2-cocycle.

6. Suppose that M is a G-module. For each $\sigma \in G$, let $m_\sigma \in M$. Show that the cochain f defined by $f(\sigma, \tau) = m_\sigma + \sigma m_\tau - m_{\sigma\tau}$ is a coboundary.

7. If M is a G-module and $f \in Z^2(G, M)$, show that $E_f = M \times G$ with multiplication defined by $(m, \sigma)(n, \tau) = (m \cdot \sigma n \cdot f(\sigma, \tau), \sigma\tau)$ makes E_f into a group.

8. If M is a G-module, show that the group extensions constructed from 2-cocycles $f, g \in Z^2(G, M)$ are isomorphic if f and g are cohomologous.

9. In the crossed product construction given in this section, show that the multiplicative identity is $f(1,1)^{-1}x_{\text{id}}$.

10. A *normalized cocycle* is a cocycle f that satisfies $f(1, \sigma) = f(\sigma, 1) = 1$ for all $\sigma \in G$. Let $A = (K/F, G, f)$ be a crossed product algebra. Show that $x_{\text{id}} = 1$ if and only if f is a normalized cocycle.

11. In the construction of group extensions, show that if e_{id} is chosen to be 1, then the resulting cocycle is a normalized cocycle.

12. Show that any 2-cocycle is cohomologous to a normalized cocycle.

13. If two crossed products $(K/F, G, f)$ and $(K/F, G, g)$ are isomorphic as F-algebras, show that f and g are cohomologous.

14. Let G be a group of order n. Show that $nH^2(G, M) = 0$. (Hint: Given f, let $c_\sigma = \sum_{\rho \in G} f(\sigma, \rho)$. Show that nf is cohomologous to the coboundary g given by $g(\sigma, \tau) = c_\sigma + \sigma c_\tau - c_{\sigma\tau}$.)

15. Let $A = (K/F, \sigma, a)$ be a cyclic algebra. If $A = \oplus_{i=0}^{n-1} Kx_{\sigma^i}$, show that $x_\sigma^n = a$.

16. *Cohomology of a cyclic group.* In this problem, we determine $H^2(G, M)$ for a cyclic group G. Suppose that $G = \langle\sigma\rangle$ is a cyclic group of order n. If M is a G-module, let $M^G = \{m \in M : \sigma m = m\}$. Also, define the *norm map* $N : M \to M^G$ by $N(m) = \sum_{i=0}^{n-1} \sigma^i m$. We will prove that $H^2(G, M) \cong M^G/\operatorname{im}(N)$ in the following steps.

(a) If $m \in M^G$, let f_m be the cochain given by $f_m(\sigma^i, \sigma^j) = 1$ if $i + j < n$, and $f_m(\sigma^i, \sigma^j) = m$ if $i + j \geq n$. Prove that f_m is a cocycle.

(b) Suppose that f_m and f_n are cocycles that are cohomologous. Then there are $c_i \in M$ with $f_m(\sigma^i, \sigma^j) = f_n(\sigma^i, \sigma^j) \cdot c_i \sigma^i(c_j) c_{i+j}^{-1}$, where we are writing c_i for c_{σ^i}. Show that $m - n = N(c_1)$.

(c) Prove that a cocycle $f \in Z^2(G, M)$ is cohomologous to f_m, where $m = \sum_{i=0}^{n-1} f(\sigma^i, \sigma)$. Make use of the cocycle condition

$$f(\sigma^i, \sigma^k) f(\sigma^{i+k}, \sigma) = \sigma^i(f(\sigma^k, \sigma)) f(\sigma^i, \sigma^{k+1}).$$

(d) Conclude from these steps that the map $m \mapsto f_m$ induces an isomorphism $M^G / \operatorname{im}(N) \cong H^2(G, M)$.

(It is known that $H^{2r}(G, M) \cong H^2(G, M)$ for a cyclic group G, so this problem calculates all of the even dimensional cohomology groups for G.)

17. In this problem, we calculate $H^1(G, M)$ for a cyclic group G. Let N be the norm map defined in the previous problem, and let $D : M \to M$ be defined by $D(m) = \sigma m - m$. We show that $H^1(G, M) \cong \ker(N) / \operatorname{im}(D)$.

(a) Let $m \in M$ satisfy $N(m) = 0$. Define a 1-cochain f by $f(\sigma^i) = m + \sigma m + \cdots + \sigma^{i-1} m$. Show that f is a 1-cocycle. For the rest of this problem, f_m will denote this cocycle.

(b) If f_m and f_n are cohomologous, show that $m - n = \sigma p - p$ for some $p \in M$.

(c) Let f be a 1-cocycle. If $m = f(\sigma)$, show that f is cohomologous to f_m.

(d) Conclude that $H^1(G, M) \cong \ker(N) / \operatorname{im}(D)$.

(Note that $H^2(G, M) \cong \ker(D) / \operatorname{im}(N)$ by the previous problem. It is known that $H^{2r+1}(G, M) \cong H^1(G, M)$ for a cyclic group G. Problems 16 and 17 then determine all of the cohomology groups for a cyclic group.)

11 Kummer Extensions

In Section 9, we described Galois extensions with cyclic Galois groups under certain restrictions on the base field. We use the results proved there together with the fundamental theorem of finite Abelian groups to characterize Galois extensions with an Abelian Galois group, provided that the base field has sufficient roots of unity.

Definition 11.1 *Let F be a field containing a primitive nth root of unity. A Galois extension K of F is called an n-Kummer extension of F provided that* $\mathrm{Gal}(K/F)$ *is an Abelian group whose exponent divides n. If K is an n-Kummer extension of F for some n, then K/F is called a Kummer extension.*

Example 11.2 If F is a field that contains a primitive nth root of unity, and if K/F is a cyclic extension of degree n, then K/F is an n-Kummer extension. If F also contains a primitive mth root of unity for some m that is a multiple of n, then K/F is also an m-Kummer extension. Therefore, if an extension is an n-Kummer extension, the integer n need not be unique.

Example 11.3 Let $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. The field K is the splitting field of $(x^2-2)(x^2-3)$ over $\mathbb{Q}$, so K is a Galois extension of $\mathbb{Q}$. A short calculation shows that $[K : \mathbb{Q}] = 4$, and the Galois group of $K/\mathbb{Q}$ consists of the four automorphisms

$$\begin{aligned} \mathrm{id} : \sqrt{2} \to \sqrt{2}, &\quad \sqrt{3} \to \sqrt{3}, \\ \sigma : \sqrt{2} \to -\sqrt{2}, &\quad \sqrt{3} \to \sqrt{3}, \\ \tau : \sqrt{2} \to \sqrt{2}, &\quad \sqrt{3} \to -\sqrt{3}, \\ \sigma\tau : \sqrt{2} \to -\sqrt{2}, &\quad \sqrt{3} \to -\sqrt{3}. \end{aligned}$$

The Galois group $\mathrm{Gal}(K/\mathbb{Q})$ is isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, an Abelian group of exponent 2. Since $\mathbb{Q}$ contains the primitive second root of unity, -1, the extension $K/\mathbb{Q}$ is a 2-Kummer extension.

The fundamental theorem of finite Abelian groups says that any such group is a direct product of cyclic groups. Using this fact together with the fundamental theorem of Galois theory and the characterization of cyclic extensions in Section 9, we obtain the following characterization of Kummer extensions.

Theorem 11.4 *Let F be a field containing a primitive nth root of unity, and let K be a finite extension of F. Then K/F is an n-Kummer extension if and only if $K = F(\sqrt[n]{a_1}, \ldots, \sqrt[n]{a_r})$ for some $a_i \in F$.*

Proof. Suppose that $K = F(\alpha_1, \ldots, \alpha_r)$ with $\alpha_i^n = a_i \in F$. If $\omega \in F$ is a primitive nth root of unity, then the distinct elements $\alpha_i, \omega\alpha_i \ldots, \omega^{n-1}\alpha_i$ are all the roots of $x^n - a_i$ in K. Thus, $x^n - a_i$ is separable over F and splits over K. Hence, K is the splitting field of the set $\{x^n - a_i : 1 \le i \le r\}$, so K/F is Galois by Theorem 4.9. If $\sigma \in \mathrm{Gal}(K/F)$, then $\sigma(\alpha_i) = \omega^j\alpha_i$ for some j since $\sigma(\alpha_i)$ is also a root of $x^n - a_i$. For each k, we see that $\sigma^k(\alpha_i) = \omega^{kj}\alpha_i$, so $\sigma^n(\alpha_i) = \alpha_i$. This is true for each i, and since the α_i generate K over F, we see that $\sigma^n = \mathrm{id}$. Therefore, the exponent of $\mathrm{Gal}(K/F)$ divides

n. To prove that $\mathrm{Gal}(K/F)$ is Abelian, take $\sigma, \tau \in \mathrm{Gal}(K/F)$. Given i, set $\sigma(\alpha_i) = \omega^j \alpha_i$ and $\tau(\alpha_i) = \omega^k \alpha_i$. Then

$$(\sigma\tau)(\alpha_i) = \sigma(\omega^k \alpha_i) = \omega^k \omega^j \alpha_i$$

and

$$(\tau\sigma)(\alpha_i) = \tau(\omega^j \alpha_i) = \omega^j \omega^k \alpha_i.$$

Thus, $\sigma\tau$ and $\tau\sigma$ agree on the generators of K, so $\sigma\tau = \tau\sigma$. In other words, $\mathrm{Gal}(K/F)$ is Abelian.

For the converse, suppose that K/F is Galois with $G = \mathrm{Gal}(K/F)$ an Abelian group whose exponent divides n. By the fundamental theorem of finite Abelian groups, $G = C_1 \times \cdots \times C_r$, where each C_i is cyclic. Note that each $|C_i|$ divides n. Let $H_i = C_1 \times \cdots \times C_{i-1} \times C_{i+1} \times \cdots \times C_r$, a subgroup of G with $G/H_i \cong C_i$. Let L_i be the fixed field of H_i. Then L_i is Galois over F, since H_i is normal in G, and $\mathrm{Gal}(L_i/F) \cong G/H_i \cong C_i$. Therefore, L_i/F is cyclic Galois. Let $[L_i : F] = m_i$. Then $m_i = |C_i|$, so m_i divides n. The field F contains the primitive m_ith root of unity ω^{n/m_i}, so by Theorem 9.5, $L_i = F(\alpha_i)$ for some $\alpha_i \in L_i$ with $\alpha_i^{m_i} \in F$. Since m_i divides n, we see that $\alpha_i^n = a_i \in F$. Under the Galois correspondence, the field $F(\alpha_1, \ldots, \alpha_r) = L_1 \cdots L_r$ corresponds to the group $H_1 \cap \cdots \cap H_r$. However, this intersection is $\langle \mathrm{id} \rangle$, so $F(\alpha_1, \ldots, \alpha_r)$ corresponds to $\langle \mathrm{id} \rangle$. Thus, $K = F(\alpha_1, \ldots, \alpha_r) = F(\sqrt[n]{a_1}, \ldots, \sqrt[n]{a_r})$. □

Example 11.5 If $K = \mathbb{Q}(\sqrt{a_1}, \ldots, \sqrt{a_r})$ for some $a_i \in \mathbb{Q}$, then $K/\mathbb{Q}$ is a 2-Kummer extension by Theorem 11.4. The degree of K/F is no larger than 2^r, but it may be less depending on the choice of the a_i. Problem 1 shows that the degree is 2^r if the a_i are distinct primes. However, $\mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{6})$ has degree 4 over $\mathbb{Q}$, not degree 8.

Example 11.6 Let $F = \mathbb{Q}(i)$, where $i = \sqrt{-1}$, and let $K = F(\sqrt[4]{12}, \sqrt[4]{3})$. Since i is a primitive fourth root of unity, K/F is a 4-Kummer extension. The degree of K/F is 8, not 16, since $K = F(\sqrt{2}, \sqrt[4]{3})$; this equality is true because $\sqrt[4]{12} = \sqrt{2}\sqrt[4]{3}$. This example shows that if $K = F(\alpha_1, \ldots, \alpha_n)$ is an n-Kummer extension of F with $\alpha_i^n \in F$, it might be the case that a smaller power of some of the α_i is also in F.

If F contains a primitive nth root of unity, then $F(\sqrt[n]{a_i}, \ldots, \sqrt[n]{a_r})$ is an n-Kummer extension of F. A basic question is to find its degree over F. Certainly, this degree is no larger than n^r. However, as the examples above show, the degree might be less than n^r. We proved in Proposition 9.6 that $[F(\sqrt[n]{a}) : F]$ is equal to the order of aF^* in the group F^*/F^{*n}. We obtain an analogous result for Kummer extensions below. However, this is a harder result, and it requires more machinery to prove. It turns out that the concept of a bilinear pairing is the right tool to investigate this question about degrees.

Definition 11.7 *Let G and H be finite Abelian groups, and let C be a cyclic group. A function $B : G \times H \to C$ is called a bilinear pairing if B is a homomorphism in each component; that is, $B(g_1g_2, h) = B(g_1, h)B(g_2, h)$ for all $g_1, g_2 \in G$ and all $h \in H$, and $B(g, h_1h_2) = B(g, h_1)B(g, h_2)$ for all $g \in G$ and all $h_1, h_2 \in H$. The pairing B is called nondegenerate if $B(g, h) = e$ for all $h \in H$ only if $g = e$, and if $B(g, h) = e$ for all $g \in G$ only if $h = e$.*

Let K/F be an n-Kummer extension, and let $\mu(F)$ be the set of all nth roots of unity in F. Then $\mu(F)$ is a cyclic group by Theorem 6.1. Also, let

$$\mathrm{KUM}(K/F) = \{a \in K^* : a^n \in F\}.$$

The set $\mathrm{KUM}(K/F)$ is a subgroup of K^*. Note that $\mathrm{KUM}(K/F)$ contains F^*, and if $K = F(\sqrt[n]{a_1}, \dots, \sqrt[n]{a_r})$, it also contains each $\sqrt[n]{a_i}$. Finally, let

$$\mathrm{kum}(K/F) = \mathrm{KUM}(K/F)/F^*.$$

We now relate bilinear pairings to Kummer extensions. We define the *Kummer pairing*

$$B : \mathrm{Gal}(K/F) \times \mathrm{kum}(K/F) \to \mu(F)$$

by

$$B(\sigma, \alpha F^*) = \sigma(\alpha)/\alpha.$$

This map is well defined, since if $\alpha F^* = \beta F^*$, then $\alpha = a\beta$ for some $a \in F^*$. Thus, $\sigma(\alpha)/\alpha = \sigma(a\beta)/a\beta = \alpha(\beta)/\beta$, since $\sigma(a) = a$.

We show that B is a nondegenerate bilinear pairing below. But first, we prove a general result about bilinear pairings that allows us to exploit the Kummer pairing to answer questions about Kummer extensions.

Lemma 11.8 *Let $B : G \times H \to C$ be a bilinear pairing. If $h \in H$, let $B_h : G \to C$ be defined by $B_h(g) = B(g, h)$. Then the map $\varphi : h \mapsto B_h$ is a group homomorphism from H to $\hom(G, C)$. If B is nondegenerate, then $\exp(G)$ divides $|C|$, the map φ is injective, and φ induces an isomorphism $G \cong H$.*

Proof. The property $B(g, h_1h_2) = B(g, h_1)B(g, h_2)$ translates to $B_{h_1h_2} = B_{h_1}B_{h_2}$. Thus, $\varphi(h_1h_2) = \varphi(h_1)\varphi(h_2)$, so φ is a homomorphism. The kernel of φ is

$$\begin{aligned}\ker(\varphi) &= \{h \in H : B_h = 0\}\\ &= \{h \in H : B(g, h) = e \text{ for all } h \in H\}.\end{aligned}$$

If φ is nondegenerate, then $\ker(\varphi) = \langle e \rangle$, so φ is injective. Suppose that $m = |C|$. Then

$$e = B(e, h) = B(g, h)^m = B(g^m, h).$$

Nondegeneracy of B forces $g^m = e$, so $\exp(G)$ divides $|G|$. By a group theory exercise (see Problems 4 and 5), $\hom(G, C)$ is isomorphic to the character group $\hom(G, \mathbb{C}^*)$, which is isomorphic to G. Therefore, there are group isomorphisms

$$H \cong \operatorname{im}(\varphi) = \hom(G, C) \cong G.$$

□

We now have the tools to investigate the Kummer pairing of a Kummer extension.

Proposition 11.9 *Let K be an n-Kummer extension of F, and let B : $\operatorname{Gal}(K/F) \times \operatorname{kum}(K/F) \to \mu(F)$ be the associated Kummer pairing. Then B is nondegenerate. Consequently,* $\operatorname{kum}(K/F) \cong \operatorname{Gal}(K/F)$.

Proof. First, we show that B is a bilinear pairing. Let $\sigma, \tau \in \operatorname{Gal}(K/F)$ and $\alpha F^* \in \operatorname{kum}(K/F)$. Then

$$\begin{aligned} B(\sigma\tau, \alpha F^*) &= \frac{\sigma\tau(\alpha)}{\alpha} = \frac{\sigma(\tau(\alpha))}{\tau(\alpha)} \cdot \frac{\tau(\alpha)}{\alpha} \\ &= \tau\left(\frac{\sigma(\alpha)}{\alpha}\right) \cdot \frac{\tau(\alpha)}{\alpha}; \end{aligned}$$

the final equality is true because $\operatorname{Gal}(K/F)$ is Abelian. But $\sigma(\alpha)^n = \alpha^n$, since $\alpha^n \in F$. Therefore, $\sigma(\alpha)/\alpha$ is an nth root of unity, so $\sigma(\alpha)/\alpha \in F$. The automorphism τ then fixes $\sigma(\alpha)/\alpha$, so

$$B(\sigma\tau, \alpha F^*) = \frac{\sigma(\alpha)}{\alpha} \cdot \frac{\tau(\alpha)}{\alpha}.$$

The pairing B is thus linear in the first component. For the second component, if $\alpha, \beta \in \operatorname{KUM}(K/F)$, then

$$B(\sigma, \alpha F^* \beta F^*) = \frac{\sigma(\alpha\beta)}{\alpha\beta} = \frac{\sigma(\alpha)\sigma(\beta)}{\alpha\beta} = \frac{\sigma(\alpha)}{\alpha} \cdot \frac{\sigma(\beta)}{\beta}.$$

Therefore, B is a bilinear pairing.

For nondegeneracy, suppose that $\sigma \in \operatorname{Gal}(K/F)$ with $B(\sigma, \alpha F^*) = 1$ for all $aF^* \in \operatorname{kum}(K/F)$. Then $\sigma(\alpha) = \alpha$ for all $\alpha \in \operatorname{KUM}(K/F)$. However, the elements in $\operatorname{KUM}(K/F)$ generate K as a field extension of F, and so automorphisms of K are determined by their action on this set. Therefore, $\sigma = \mathrm{id}$. Also, if $B(\sigma, \alpha F^*) = 1$ for all $\sigma \in \operatorname{Gal}(K/F)$, then $\sigma(\alpha) = \alpha$ for all σ. But then $\alpha \in \mathcal{F}(\operatorname{Gal}(K/F))$, and this fixed field is F by the fundamental theorem. Therefore, $\alpha F^* = F^*$, so B is nondegenerate. The isomorphism $\operatorname{kum}(K/F) \cong \operatorname{Gal}(K/F)$ then follows from Lemma 11.8. □

If K/F is a Galois extension, then $[K : F] = |\mathrm{Gal}(K/F)|$. If, in addition, K is a Kummer extension of F, then Proposition 11.9 shows that $[K : F] = |\mathrm{kum}(K/F)|$. Therefore, if we can determine $\mathrm{kum}(K/F)$, then among other things we know the degree of K/F. The following result is a generalization of Theorem 9.6.

Proposition 11.10 *Let K/F be an n-Kummer extension. Then there is an injective group homomorphism $f : \mathrm{kum}(K/F) \to F^*/F^{*n}$, given by $f(\alpha F^*) = \alpha^n F^{*n}$. The image of f is then a finite subgroup of F^*/F^{*n} of order equal to $[K : F]$.*

Proof. It is easy to see that f is well defined and that f preserves multiplication. For injectivity, let $\alpha F^* \in \ker(f)$. Then $\alpha^n \in F^{*n}$, so $\alpha^n = a^n$ for some $a \in F$. Hence, α/a is an nth root of unity, and so $\alpha/a \in F$. Therefore, $\alpha \in F$, so $\alpha F^* = F^*$ is the identity. The group $\mathrm{kum}(K/F)$ is then isomorphic to the image of f. The final statement of the proposition follows immediately from Proposition 11.9. □

This proposition can be used in reverse to construct Kummer extensions of a given degree. Let G be a finite Abelian subgroup of F^*/F^{*n}. In a fixed algebraic closure of F, let

$$F(G) = (\{F\sqrt[n]{a} : aF^{*n} \in G\}).$$

Problem 6 shows that $F(G)$ is an n-Kummer extension with Galois group $\mathrm{Gal}(F(G)/F) \cong G$, and so $[F(G) : F] = |G|$.

Example 11.11 Let $F = \mathbb{C}(x, y, z)$ be the rational function field in three variables over $\mathbb{C}$, and let $K = F(\sqrt[4]{xyz}, \sqrt[4]{y^2z}, \sqrt[4]{xz^2})$. Then K/F is a 4-Kummer extension. The image of $\mathrm{kum}(K/F)$ in F^*/F^{*4} is generated by the cosets of xyz, yz, and xz^2. For simplicity we will call these three cosets a, b, c respectively. We claim that the subgroup of F^*/F^{*4} generated by a, b, c has order 32, which shows that $[K : F] = 32$ by Proposition 11.10. The subgroup $\langle a, b\rangle$ of F^*/F^{*4} generated by a and b has order 16, since the 16 elements a^ib^j with $1 \le i, j \le 4$ are all distinct. To see this, suppose that $a^ib^j = a^kb^l$. Then there is an $h \in F^*$ with

$$(xyz)^i(y^2z)^j = (xyz)^k(y^2z)^l h^4.$$

Writing $h = f/g$ with $f, g \in \mathbb{C}[x, y, z]$ relatively prime gives

$$(xyz)^i(y^2z)^j f(x, y, z) = (xyz)^k(y^2z)^l g(x, y, z)^4.$$

By unique factorization, comparing powers of x and z on both sides of this equation, we obtain

$$i \equiv k \pmod 4,$$
$$i + j \equiv k + l \pmod 4.$$

These equations force $i \equiv k \pmod 4$ and $j \equiv l \pmod 4$, so the elements $a^i b^j$ for $1 \le i, j \le 4$ are indeed distinct. Note that $abc = x^2y^2z^4F^{*4}$, so $(abc)^2 = x^4y^4z^8F^{*4} = F^{*4}$. Therefore, $c^2 = (ab)^2$, so either the subgroup $\langle a, b, c\rangle$ of F^*/F^{*4} generated by a, b, c is equal to $\langle a, b\rangle$, or $\langle a, b\rangle$ has index 2 in $\langle a, b, c\rangle$. For the first to happen, we must have $c = a^i b^j$ for some i, j. This leads to an equation

$$xz^2 f(x,y,z)^4 = (xyz)^i (y^2 z)^j g(x,y,z)^4$$

for some polynomials f, g. Again applying unique factorization and equating powers of x and y gives $1 \equiv i \pmod 4$ and $0 \equiv i + 2j \pmod 4$. A simultaneous solution of these equations does not exist, so c is not in the group $\langle a, b\rangle$, so $\langle a, b\rangle$ has index 2 in $\langle a, b, c\rangle$. This proves that $\langle a, b, c\rangle$ has order 32, as we wanted to show.

Problems

1. Let $p_1, \ldots, p_n$ be distinct primes. Show that $[\mathbb{Q}(\sqrt{p_1}, \ldots, \sqrt{p_n}) : \mathbb{Q}] = 2^n$.

2. Let $F = \mathbb{Q}(\{\sqrt{n} : 1 \le n \le 28\})$. Determine $[F : \mathbb{Q}]$.

3. Let N be a positive integer, and let $F_N = \mathbb{Q}(\{\sqrt{n} : 1 \le n \le N\})$. Determine $[F_N : \mathbb{Q}]$.

4. Let G be a finite Abelian group whose exponent divides the order of a cyclic group C. Show that $\hom(G, C) \cong \hom(G, \mathbb{C})$.

5. If G is a finite Abelian group, show that $\hom(G, \mathbb{C}) \cong G$. (Hint: First prove this if G is cyclic, then show that $\hom(G_1 \times G_2, \mathbb{C}) \cong \hom(G_1, \mathbb{C}) \times \hom(G_2, \mathbb{C})$, and then invoke the structure theorem for finite Abelian groups.)

6. Let F be a field containing a primitive nth root of unity, and let G be a subgroup of F^*/F^{*n}. Let $F(G) = F(\{\sqrt[n]{a} : aF^{*n} \in G\})$. Show that $F(G)$ is an n-Kummer extension of F and that G is the image of $\operatorname{kum}(F(G)/F)$ under the map f defined in Proposition 11.10. Conclude that $\operatorname{Gal}(K/F) \cong G$ and $[F(G) : F] = |G|$.

III
Applications of Galois Theory

Now that we have developed Galois theory and have investigated a number of types of field extensions, we can put our knowledge to use to answer some of the most famous questions in mathematical history. In Section 15, we look at ruler and compass constructions and prove that with ruler and compass alone it is impossible to trisect an arbitrary angle, to duplicate the cube, to square the circle, and to construct most regular n-gons. These questions arose in the days of the ancient Greeks but were left unanswered for 2500 years. In order to prove that it is impossible to square the circle, we prove in Section 14 that π is transcendental over $\mathbb{Q}$, and we prove at the same time that e is also transcendental over $\mathbb{Q}$. In Section 16, we prove that there is no algebraic formula, involving only field operations and extraction of roots, to find the roots of an arbitrary nth degree polynomial if $n \geq 5$. Before doing so, we investigate in detail polynomials of degree less than 5. By the mid-sixteenth century, formulas for finding the roots of quadratic, cubic, and quartic polynomials had been found. The success in finding the roots of arbitrary cubics and quartics within a few years of each other led people to believe that formulas for arbitrary degree polynomials would be found. However, it was not until the early nineteenth century that Abel was able to prove that it is impossible to find an algebraic formula for the roots of an arbitrary fifth degree polynomial, and Galois was able to use his new theory to explain why some polynomials had formulas for their roots and others did not.

12 Discriminants

In this section, we define discriminants and give methods to calculate them. The discriminant of a polynomial is a generalization to arbitrary degree polynomials of the discriminant of a quadratic. If $K = F(a)$ is a Galois extension of a field F, and if $f = \min(F, a)$, then the Galois group $\mathrm{Gal}(K/F)$ can be viewed as a subgroup of the group of permutations of the roots of f. The discriminant determines when this subgroup consists solely of even permutations. We will use this information to describe the splitting field of a polynomial of degree 4 or less in Section 13. While we only need a little information about discriminants in Section 13, we go into some detail here for two reasons. First, there are some interesting relations that make calculating discriminants manageable, and there are notions of discriminants in a number of other places, such as algebraic number theory, quadratic form theory, and noncommutative ring theory. While the different notions of discriminant may seem unrelated, this is not the case, as we point out in the following discussion.

The discriminant of a polynomial and an element

The type of discriminant we need in Section 13 is the discriminant of a polynomial. To motivate the definition, consider a quadratic polynomial $f(x) = x^2 + bx + c$ whose discriminant is $b^2 - 4c$. The roots of f are $\alpha_1 = \frac{1}{2}(-b + \sqrt{b^2 - 4c})$ and $\alpha_2 = \frac{1}{2}(-b - \sqrt{b^2 - 4c})$. Therefore, $\sqrt{b^2 - 4c} = \alpha_1 - \alpha_2$, so $b^2 - 4c = (\alpha_1 - \alpha_2)^2$. This indicates a way to generalize the notion of the discriminant of a quadratic to higher degree polynomials.

Definition 12.1 *Let F be a field with* $\mathrm{char}(F) \neq 2$, *and let $f(x) \in F[x]$. Let $\alpha_1, \ldots, \alpha_n$ be the roots of f in some splitting field K of f over F, and let $\Delta = \prod_{i<j}(\alpha_i - \alpha_j) \in K$. Then the discriminant* $\mathrm{disc}(f)$ *of f is the element $D = \Delta^2 = \prod_{i<j}(\alpha_i - \alpha_j)^2$.*

Definition 12.2 *If K is an algebraic extension of F with* $\mathrm{char}(F) \neq 2$ *and $\alpha \in K$, then the discriminant* $\mathrm{disc}(\alpha)$ *is* $\mathrm{disc}(\min(F, \alpha))$.

The discriminant $\mathrm{disc}(\alpha)$ defined above is dependent on the base field F. Also, the element Δ is dependent on the labeling of the roots of f, in that a different labeling can change Δ by -1. However, the discriminant does not depend on this labeling. Note that if $f(x) \in F[x]$, then $D = \mathrm{disc}(f) = 0$ if and only if f has a repeated root. The discriminant thus will give us information only when f has no repeated roots. It is in this case that we concentrate our investigation. The discriminant D clearly is an element of K. We can say more than that. If K is the splitting field of a separable, irreducible polynomial $f \in F[x]$ of degree n over F, then we view $\mathrm{Gal}(K/F)$ as a subgroup of S_n by viewing the elements of $\mathrm{Gal}(K/F)$ as permutations of the roots of f.

Lemma 12.3 *Let F be a field with* $\mathrm{char}(F) \neq 2$, *let $f(x) \in F[x]$ be an irreducible, separable polynomial, and let K be the splitting field of $f(x)$ over F. If Δ is defined as in Definition* 12.2, *then $\sigma \in \mathrm{Gal}(K/F)$ is an even permutation if and only if $\sigma(\Delta) = \Delta$, and σ is odd if and only if $\sigma(\Delta) = -\Delta$. Furthermore,* $\mathrm{disc}(f) \in F$.

Proof. Before we prove this, we note that the proof we give is the same as the typical proof that every permutation of S_n is either even or odd. In fact, the proof of this result about S_n is really about discriminants. It is easy to see that each $\sigma \in G = \mathrm{Gal}(K/F)$ fixes $\mathrm{disc}(f)$, so $\mathrm{disc}(f) \in F$. For the proof of the first statement, if $n = \deg(f)$, let $M = F(x_1, \ldots, x_n)$. We saw in Example 2.22 that S_n acts as field automorphisms on M by permuting the variables. Let $h(x) = \prod_{i<j}(x_i - x_j)$. Suppose that $\sigma \in S_n$ is a transposition, say $\sigma = (ij)$ with $i < j$. Then σ affects only those factors of h that involve i or j. We break up these factors into four groups:

$$\begin{array}{rl} x_i - x_j & \\ x_k - x_i,\ x_k - x_j & \text{for} \quad k < i, \\ x_i - x_l,\ x_j - x_l & \text{for} \quad j < l, \\ x_i - x_m,\ x_m - x_j & \text{for} \quad i < m < j. \end{array}$$

For $k < i$, the permutation $\sigma = (ij)$ maps $x_k - x_i$ to $x_k - x_j$ and vice versa, and σ maps $x_i - x_l$ to $x_j - x_l$ and vice versa for $j < l$. If $i < m < j$, then

$$\sigma(x_i - x_m) = x_j - x_m = -(x_m - x_j)$$

and

$$\sigma(x_m - x_j) = x_m - x_i = -(x_i - x_m).$$

Finally,

$$\sigma(x_i - x_j) = x_j - x_i = -(x_i - x_j).$$

Multiplying all the terms together gives $\sigma(h) = -h$. Thus, we see for an arbitrary $\sigma \in S_n$ that $\sigma(h) = h$ if and only if σ is a product of an even number of permutations, and $\sigma(h) = -h$ if and only if σ is a product of an odd number of permutations. By substituting the roots α_i of f for the x_i, we obtain the desired conclusion. □

Recall that the set A_n of all even permutations in S_n is a subgroup; it is called the *alternating group.*

Corollary 12.4 *Let F, K, and f be as in Lemma* 12.3, *and let $G = \mathrm{Gal}(K/F)$. Then $G \subseteq A_n$ if and only if* $\mathrm{disc}(f) \in F^2$. *Under the correspondence of the fundamental theorem, the field $F(\Delta) \subseteq K$ corresponds to the subgroup $G \cap A_n$ of G.*

Proof. This follows from the lemma, since $G \subseteq A_n$ if and only if each $\sigma \in G$ is even, and this occurs if and only if $\sigma(\Delta) = \Delta$. Therefore, $G \subseteq A_n$ if and only if $\operatorname{disc}(f) \in F^2$. □

One problem with the definition of a discriminant is that in order to calculate it we need the roots of the polynomial. We will give other descriptions of the discriminant that do not require knowledge of the roots and lend themselves to calculation. We first obtain a description of the discriminant in terms of determinants.

Let K be a field and let $\alpha_1, \ldots, \alpha_n \in K$. Then the *Vandermonde matrix* $V(\alpha_1, \ldots, \alpha_n)$ is the $n \times n$ matrix

$$V(\alpha_1, \ldots, \alpha_n) = \begin{bmatrix} 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_n & \alpha_n^2 & \cdots & \alpha_n^{n-1} \end{bmatrix}.$$

Lemma 12.5 *If K is a field and $\alpha_1, \ldots, \alpha_n \in K$, then the determinant of the Vandermonde matrix $V(\alpha_1, \ldots, \alpha_n)$ is $\prod_{i<j}(\alpha_j - \alpha_i)$. Consequently, if $f \in F[x]$ has roots $\alpha_1, \ldots, \alpha_n \in K$ in some extension K of F, then the discriminant of f is equal to $(\det(V(\alpha_1, \ldots, \alpha_n)))^2$.*

Proof. Let $A = V(\alpha_1, \ldots, \alpha_n)$. That $\det(A) = \prod_{i<j}(\alpha_j - \alpha_i)$ is a moderately standard fact from linear algebra. For those who have not seen this, we give a proof. Note that if $\alpha_i = \alpha_j$ with $i \neq j$, then $\det(A) = 0$, since two rows of A are the same, so the determinant formula is true in this case. We therefore assume that the α_i are distinct, and we prove the result using induction on n. If $n = 1$, this is clear, so suppose that $n > 1$. Let $h(x) = \det(V(\alpha_1, \alpha_2, \ldots, \alpha_{n-1}, x))$. Then $h(x)$ is a polynomial of degree less than n. By expanding the determinant about the last row, we see that the leading coefficient of h is $\det(V(\alpha_1, \alpha_2, \ldots, \alpha_{n-1}))$. Moreover, $h(\alpha_i) = \det(V(\alpha_1, \ldots, \alpha_{n-1}, \alpha_i))$, so $h(\alpha_i) = 0$ if $1 \leq i \leq n-1$. Therefore, $h(x)$ is divisible by each $x - \alpha_i$. Since $\deg(h) < n$ and h has $n-1$ distinct factors, $h(x) = c(x-\alpha_1)\cdots(x-\alpha_{n-1})$, where $c = \det(V(\alpha_1, \alpha_2, \ldots, \alpha_{n-1}))$. By evaluating h at α_n and using induction, we get

$$\begin{aligned} h(\alpha_n) &= \det(V(\alpha_1, \alpha_2, \ldots, \alpha_n)) \\ &= \prod_{i<j\leq n-1} (\alpha_j - \alpha_i) \prod_{i<n} (\alpha_n - \alpha_i) \\ &= \prod_{i<j} (\alpha_j - \alpha_i). \end{aligned}$$

This finishes the proof that $\det(V(\alpha_1, \alpha_2, \ldots, \alpha_n)) = \prod_{i<j}(\alpha_j - \alpha_i)$. The last statement of the lemma is an immediate consequence of this formula and the definition of discriminant. □

The discriminant of a polynomial can be determined by the coefficients without having to find the roots, as we proceed to show. This is a convenient fact and will be used in Section 13 to describe polynomials of degree 3 and 4. Let $A = V(\alpha_1, \dots, \alpha_n)$. Then $\det(A)^2 = \det(A^t A)$. Moreover,

$$A^t A = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{n-1} & \alpha_2^{n-1} & \cdots & \alpha_n^{n-1} \end{bmatrix} \cdot \begin{bmatrix} 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_n & \alpha_n^2 & \cdots & \alpha_n^{n-1} \end{bmatrix}$$
$$= \begin{bmatrix} t_0 & t_1 & \cdots & t_{n-1} \\ t_1 & t_2 & \cdots & t_n \\ \vdots & \vdots & \ddots & \vdots \\ t_{n-1} & t_n & \cdots & t_{2n-2} \end{bmatrix},$$

where $t_i = \sum_j \alpha_j^i$ for $i \geq 1$, and $t_0 = n$. Therefore, $\det(A)^2$ is the determinant of this latter matrix. This is helpful because if the roots of $f(x)$ are $\alpha_1, \dots, \alpha_n$, then there are recursive relations between the t_i and the coefficients of f, and so the determinant of the t_i can be found in terms of the coefficients of f. These relations are called *Newton's identities.* Note that $t_i = T_{K/F}(\alpha_1^i)$ if K is the splitting field of $\min(F, \alpha_1)$.

Proposition 12.6 (Newton's Identities) *Let $f(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n$ be a monic polynomial over F with roots $\alpha_1, \dots, \alpha_n$. If $t_i = \sum_j \alpha_j^i$, then*

$$t_m + a_{n-1}t_{m-1} + \cdots + a_{n-m+1}t_1 + m a_{n-m} = 0 \quad \textit{for } m \leq n,$$
$$t_m + a_{n-1}t_{m-1} + \cdots + a_0 t_{m-n} = 0 \quad \textit{for } m > n.$$

Proof. An alternative way of stating Newton's identities is to use the elementary symmetric functions s_i in the a_i, instead of the a_i. Since $s_i = (-1)^i a_{n-i}$, Newton's identities can also be written as

$$t_m - s_1 t_{m-1} + s_2 t_{m-2} - + \cdots (-1)^m m s_m = 0 \quad \text{for } m \leq n$$
$$t_m - a_1 t_{m-1} - + \cdots + (-1)^n s_n t_{m-n} = 0 \quad \text{for } m > n.$$

The proof we give here is from Mead [21]. The key is arranging the terms in the identities in a useful manner. We start with a bit of notation. If $(a_1, a_2, \dots, a_r)$ is a sequence of nonincreasing, nonnegative integers, let

$$f_{(a_1, a_2, \dots, a_r)} = \sum \alpha_{\sigma(1)}^{a_1} \cdots \alpha_{\sigma(n)}^{a_r},$$

where the sum is over all permutations σ of $\{1, 2, \dots, n\}$ that give distinct terms. Then $s_i = f_{(1,1,\dots,1)}$ (i ones) and $t_i = f_{(i)}$. To simplify the notation a little, the sequence of i ones will be denoted (1_i), and the sequence

$(a, 1, \ldots, 1)$ of length $i+1$ will be denoted $(a, 1_i)$. It is then straightforward to see that

$$\begin{aligned} f_{(m-1)}f_{(1)} &= f_{(m)} + f_{(m-1,1)}, \\ f_{(m-2)}f_{(1,1)} &= f_{(m-1,1)} + f_{(m-2,1,1)}, \\ f_{(m-3)}f_{(1,1,1)} &= f_{(m-2,1,1)} + f_{(m-3,1,1,1)}, \end{aligned}$$

and, in general,

$$f_{(m-i)}f_{(1_i)} = f_{(m-i+1,1_i)} + f_{(m-i,1_i)} \quad \text{for } 1 \le i < \min\{m-1, n\}. \quad (12.1)$$

Moreover, if $m \le n$ and $i = m-1$, then

$$f_{(1)}f_{(1_{m-1})} = f_{(2,1_{m-2})} + mf_{(1_m)}.$$

If $m > n = i$, then

$$f_{(m-n)}f_{(1_n)} = f_{(m-n+1,1_{n-1})}.$$

Newton's identities then follow from these equations by multiplying the ith equation in (12.1) by $(-1)^{i-1}$ and summing over i. □

Newton's identities together with Lemma 12.5 give us a manageable way of calculating discriminants of polynomials. As an illustration, we determine the discriminant of a quadratic and of a cubic. The calculation of the discriminant of a cubic will come up in Section 13.

Example 12.7 Let $f(x) = x^2 + bx + c$. Then $t_0 = 2$. Also, Newton's identities yield $t_1 + b = 0$, so $t_1 = -b$. For t_2, we have $t_2 + bt_1 + 2c = 0$, so $t_2 = -bt_1 - 2c = b^2 - 2c$. Therefore,

$$\operatorname{disc}(f) = \begin{vmatrix} 2 & -b \\ -b & b^2 - 2c \end{vmatrix} = 2(b^2 - 2c) - b^2 = b^2 - 4c,$$

the usual discriminant of a monic quadratic.

Example 12.8 Let $f(x) = x^3 + px + q$. Then $a_0 = q$, $a_1 = p$, and $a_2 = 0$, so by Newton's identities we get

$$\begin{aligned} t_1 &= 0, \\ t_2 &= -2p, \\ t_3 &= -3q, \\ t_4 &= 2p^2. \end{aligned}$$

Therefore

$$\operatorname{disc}(f) = \begin{vmatrix} 3 & 0 & -2p \\ 0 & -2p & -3q \\ -2p & -3q & 2p^2 \end{vmatrix} = -4p^3 - 27q^2.$$

For an arbitrary monic cubic, we could do a similar calculation, but looking ahead to Section 13, where we find the roots of a cubic, we note that the case above is sufficient. For, if $g(x) = x^3 + ax^2 + bx + c$, let $y = x - a/3$. By Taylor expansion, we have

$$g(x) = g(a/3) + g'(a/3)(x - a/3) + \frac{g''(a/3)}{2!}(x - a/3)^2 + \frac{g'''(a/3)}{3!}(x - a/3)^3.$$

The choice of y was made to satisfy $g''(a/3) = 0$. If $p = g'(a/3)$ and $q = g(a/3)$, then $g(x) = y^3 + py + q$. If the roots of g are α_1, α_2, and α_3, then the roots of $y^3 + py + q$ are $\alpha_1 - a/3$, $\alpha_2 - a/3$, and $\alpha_3 - a/3$. Therefore, the definition of discriminant shows that $\operatorname{disc}(g(x)) = \operatorname{disc}(y^3 + py + q)$. The interested reader can check that $\operatorname{disc}(g(x)) = a^2(b^2 - 4ac) - 4b^3 - 27c^2 + 18abc$.

We give a further description of the discriminant, this time in terms of norms.

Proposition 12.9 *Let $L = F(\alpha)$ be a field extension of F. If $f(x) = \min(F, \alpha)$, then $\operatorname{disc}(f) = (-1)^{n(n-1)/2} N_{L/F}(f'(\alpha))$, where $f'(x)$ is the formal derivative of f.*

Proof. Let K be a splitting field for f over F, and write $f(x) = (x - \alpha_1) \cdots (x - \alpha_n) \in K[x]$. Set $\alpha = \alpha_1$. Then a short calculation shows that $f'(\alpha_j) = \prod_{i=1, i \neq j}^{n} (\alpha_j - \alpha_i)$. If $\sigma_1, \ldots, \sigma_n$ are the F-homomorphisms of L to K that satisfy $\sigma_i(\alpha) = \alpha_i$, then by Proposition 8.12,

$$N_{L/F}(f'(\alpha)) = \prod_j \sigma_j(f'(\alpha)) = \prod_j f'(\alpha_j).$$

Using the formula above for $f'(\alpha_j)$, we see by checking signs carefully that

$$N_{L/F}(f'(\alpha)) = \prod_j f'(\alpha_j) = \prod_j \prod_{\substack{i=1 \\ i \neq j}}^{n} (\alpha_j - \alpha_i) = (-1)^{n(n-1)/2} \operatorname{disc}(f).$$

□

Example 12.10 Let p be an odd prime, and let ω be a primitive pth root of unity in $\mathbb{C}$. We use the previous result to determine $\operatorname{disc}(\omega)$. Let $K = \mathbb{Q}(\omega)$, the pth cyclotomic extension of $\mathbb{Q}$. If $f(x) = \min(\mathbb{Q}, \omega)$, then $f(x) = 1 + x + \cdots + x^{p-1} = (x^p - 1)/(x - 1)$. We need to calculate $N_{K/\mathbb{Q}}(f'(\omega))$. First,

$$f'(x) = \frac{px^{p-1}(x-1) - (x^p - 1)}{(x-1)^2},$$

so $f'(\omega) = p\omega^{p-1}/(\omega - 1)$. We claim that $N_{K/\mathbb{Q}}(\omega) = 1$ and $N_{K/\mathbb{Q}}(\omega - 1) = p$. To prove the first equality, by the description of $\operatorname{Gal}(K/\mathbb{Q})$ given in

Corollary 7.8, we have

$$N_{K/\mathbb{Q}}(\omega) = \prod_{i=1}^{p-1} \omega^i = \omega^{p(p-1)/2} = 1$$

since p is odd. For the second equality, note that

$$1 + x + \cdots + x^{p-1} = \prod_{i=1}^{p-1} (x - \omega^i),$$

so $p = \prod_{i=1}^{p-1}(1 - \omega^i)$. However,

$$N_{K/\mathbb{Q}}(\omega - 1) = \prod_{i=1}^{p-1} (\omega^i - 1),$$

so $N_{K/\mathbb{Q}}(\omega - 1) = p$, where again we use p odd. From this, we see that

$$\begin{aligned} N_{K/\mathbb{Q}}(f'(\omega)) &= N_{K/\mathbb{Q}}\left(\frac{p\omega^{p-1}}{\omega - 1}\right) = \frac{N_{K/\mathbb{Q}}(p) N_{K/\mathbb{Q}}(\omega)^{p-1}}{N_{K/\mathbb{Q}}(\omega - 1)} \\ &= \frac{p^{p-1} \cdot 1}{p} = p^{p-2}. \end{aligned}$$

The discriminant of an n-tuple and of a field extension

We now define the discriminant of a field extension of degree n and of an n-tuple in the field extension. We shall see that our definition of the discriminant of an element is a special case of this new definition. Let K be a separable extension of F with $[K : F] = n$. Recall from Lemma 8.9 that $[K : F]$ is equal to the number of F-homomorphisms from K into an algebraic closure of F.

Definition 12.11 *Let K be a separable extension of F of degree n, and let $\sigma_1, \sigma_2, \ldots, \sigma_n$ be the distinct F-homomorphisms from K to an algebraic closure of F. If $\alpha_1, \alpha_2, \ldots, \alpha_n$ are any n elements of K, then the discriminant of the n-tuple $(\alpha_1, \ldots, \alpha_n)$ is $\operatorname{disc}(\alpha_1, \ldots, \alpha_n) = \det(\sigma_i(\alpha_j))^2$. If $\beta_1, \ldots, \beta_n$ is any F-basis of K, then the discriminant of the field extension K/F is $\operatorname{disc}(K/F) = \operatorname{disc}(\beta_1, \ldots, \beta_n)$.*

The definition of $\operatorname{disc}(K/F)$ depends on the choice of basis. We will show just how it depends on the basis. But first, we give another description of the discriminant of an n-tuple, which will show us that this discriminant is an element of the base field F.

Lemma 12.12 *Let K be a separable field extension of F of degree n, and let $\alpha_1, \ldots, \alpha_n \in K$. Then $\operatorname{disc}(\alpha_1, \ldots, \alpha_n) = \det(\operatorname{Tr}_{K/F}(\alpha_i \alpha_j))$. Consequently, $\operatorname{disc}(\alpha_1, \ldots, \alpha_n) \in F$.*

Proof. Let $\sigma_1, \ldots, \sigma_n$ be the distinct F-homomorphisms from K to an algebraic closure of F. If $A = (\sigma_i(\alpha_j))$, then the discriminant of the n-tuple $\alpha_1, \ldots, \alpha_n$ is the determinant of the matrix $A^t A$, whose ij entry is

$$\begin{aligned}\sum_k \sigma_k(\alpha_i)\sigma_k(\alpha_j) &= \sum_k \sigma_k(\alpha_i\alpha_j) \\ &= \mathrm{Tr}_{K/F}(\alpha_i\alpha_j).\end{aligned}$$

Therefore, $\mathrm{disc}(\alpha_1, \ldots, \alpha_n) = \det(\mathrm{Tr}_{K/F}(\alpha_i\alpha_j))$. □

The next result shows that the discriminant can be used to test whether or not an n-tuple in K forms a basis for K.

Proposition 12.13 *Let K be a separable field extension of F of degree n, and let $\alpha_1, \ldots, \alpha_n \in K$. Then $\mathrm{disc}(\alpha_1, \ldots, \alpha_n) = 0$ if and only if $\alpha_1, \ldots, \alpha_n$ are linearly dependent over F. Thus, $\{\alpha_1, \ldots, \alpha_n\}$ is an F-basis for K if and only if $\mathrm{disc}(\alpha_1, \ldots, \alpha_n) \neq 0$.*

Proof. Suppose that the α_i are linearly dependent over F. Then one of the α_i is an F-linear combination of the others. If $\alpha_i = \sum_{k \neq i} a_k\alpha_k$ with $a_j \in F$, then

$$\mathrm{Tr}_{K/F}(\alpha_i\alpha_j) = \sum_k a_k \, \mathrm{Tr}_{K/F}(\alpha_k\alpha_j).$$

Therefore, the columns of the matrix $(\mathrm{Tr}_{K/F}(\alpha_i\alpha_j))$ are linearly dependent over F, so $\det(\mathrm{Tr}_{K/F}(\alpha_i\alpha_j)) = 0$.

Conversely, suppose that $\det(\mathrm{Tr}_{K/F}(\alpha_i\alpha_j)) = 0$. Then the rows $R_1, \ldots, R_n$ of the matrix $(\mathrm{Tr}_{K/F}(\alpha_i\alpha_j))$ are dependent over F, so there are $a_i \in F$, not all zero, with $\sum_i a_i R_i = 0$. The vector equation $\sum_i a_i R_i = 0$ means that $\sum_i a_i \, \mathrm{Tr}_{K/F}(\alpha_i\alpha_j) = 0$ for each j. Let $x = \sum_i a_i\alpha_i$. By linearity of the trace, we see that $\mathrm{Tr}_{K/F}(x\alpha_j) = 0$ for each j. If the α_i are independent over F, then they form a basis for K. Consequently, linearity of the trace then implies that $\mathrm{Tr}_{K/F}(xy) = 0$ for all $y \in K$. This means that the trace map is identically zero, which is false by the Dedekind independence lemma. Thus, the α_i are dependent over F. □

We now see exactly how the discriminant of a field extension depends on the basis chosen to calculate it.

Proposition 12.14 *Let $\{\alpha_1, \ldots, \alpha_n\}$ and $\{\beta_1, \ldots, \beta_n\}$ be two F-bases for K. Let $A = (a_{ij})$ be the $n \times n$ transition matrix between the two bases; that is, $\beta_j = \sum_i a_{ij}\alpha_i$. Then $\mathrm{disc}(\beta_1, \ldots, \beta_n) = \det(A)^2 \, \mathrm{disc}(\alpha_1, \ldots, \alpha_n)$. Consequently, the coset of $\mathrm{disc}(K/F)$ in F^*/F^{*2} is well defined, independent of the basis chosen.*

Proof. Since $\beta_j = \sum_k a_{kj}\alpha_k$, we have $\sigma_i(\beta_j) = \sum_k a_{kj}\sigma_i(\alpha_k)$. In terms of matrices, this says that

$$(\sigma_i(\beta_j)) = (a_{ij})^t(\sigma_i(\alpha_j)) = A^t(\sigma_i(\alpha_j)).$$

Therefore, by taking determinants, we obtain

$$\operatorname{disc}(\beta_1, \ldots, \beta_n) = \det(A)^2 \operatorname{disc}(\alpha_1, \ldots, \alpha_n).$$

The final statement of the proposition follows immediately from this relation, together with the fact that the discriminant of a basis is nonzero, by Proposition 12.13. □

To make the definition of discriminant of a field extension well defined, one can define it to be the coset in F^*/F^{*2} represented by $\operatorname{disc}(\alpha_1, \ldots, \alpha_n)$ for any basis $\{\alpha_1, \ldots, \alpha_n\}$ of K. This eliminates ambiguity, although it is not always the most convenient way to work with discriminants.

Example 12.15 In this example, we show that the discriminant of a polynomial is equal to the discriminant of an appropriate field extension. Suppose that $K = F(\alpha)$ is an extension of F of degree n. Then $1, \alpha, \alpha^2, \ldots, \alpha^{n-1}$ is a basis for K. We calculate $\operatorname{disc}(K/F)$ relative to this basis. We have $\operatorname{disc}(K/F) = \det(\sigma_i(\alpha^{j-1}))^2$. Consequently, if $\alpha_i = \sigma_i(\alpha)$, then

$$\begin{aligned}\operatorname{disc}(K/F) &= \det\begin{pmatrix} 1 & \sigma_1(\alpha) & \cdots & \sigma_1(\alpha^{n-1}) \\ 1 & \sigma_2(\alpha) & \cdots & \sigma_2(\alpha^{n-1}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \sigma_n(\alpha) & \cdots & \sigma_n(\alpha^{n-1}) \end{pmatrix}^2 \\ &= \det(V(\alpha_1, \alpha_2, \ldots, \alpha_n))^2.\end{aligned}$$

Therefore, $\operatorname{disc}(K/F) = \operatorname{disc}(\alpha) = \operatorname{disc}(\min(F, \alpha))$.

Example 12.16 Let $K = \mathbb{Q}(\sqrt{-1})$. If $i = \sqrt{-1}$, then using the basis $1, i$ of $K/\mathbb{Q}$, we get

$$\operatorname{disc}(\mathbb{Q}(i)/\mathbb{Q}) = \det\begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix}^2 = (-2i)^2 = -4.$$

More generally, if $K = \mathbb{Q}(\sqrt{d})$ with d a square-free integer, then using $1, \sqrt{d}$ as a basis, we see that the discriminant is $4d$.

The discriminant of a bilinear form

We now extend the idea of discriminant to its most general form that we consider. The two previous notions of discriminant will be special cases of

this general form. The starting point here is similar to that considered in Section 11, when we discussed Kummer pairings. If V is an F-vector space, a *bilinear form* on V is a mapping $B : V \times V \to F$ that is linear in each variable. In other words, for all $u, v, w \in V$ and all $\alpha, \beta \in F$, we have

$$B(u, \alpha v + \beta w) = \alpha B(u, v) + \beta B(u, w),$$
$$B(\alpha u + \beta v, w) = \alpha B(u, w) + \beta B(v, w).$$

Definition 12.17 *If V is an F-vector space and if $B : V \times V \to F$ is a bilinear form, then the discriminant of B relative to a basis $\mathcal{V} = \{v_1, \ldots, v_n\}$ of V is* $\operatorname{disc}(B)_{\mathcal{V}} = \det(B(v_i, v_j))$.

As with the discriminant of a field extension, this definition depends on the choice of basis. If $\mathcal{W} = \{w_1, \ldots, w_n\}$ is another basis, let A be the matrix describing the basis change; that is, if $A = (a_{ij})$, then $w_j = \sum_i a_{ij} v_i$. By the bilinearity of B, we have

$$B(w_i, w_j) = B\left(\sum_k a_{ik} v_k, \sum_l a_{jl} v_l\right) = \sum_{k,l} a_{ik} B(v_k, v_l) a_{jl}.$$

Therefore, it follows that $(B(w_i, w_j)) = A^t(B(v_k, v_l))A$. Taking determinants gives

$$\operatorname{disc}(B)_{\mathcal{W}} = \det(A)^2 \operatorname{disc}(B)_{\mathcal{V}},$$

the same relation that was found for field extensions.

A bilinear form is *nondegenerate* if $B(v, w) = 0$ for all w only if $v = 0$, and if $B(v, w) = 0$ for all v only if $w = 0$. As in Section 11, if we define $B_v : V \to F$ by $B_v(w) = B(v, w)$, then the map $v \mapsto B_v$ is a homomorphism from V to $\hom_F(V, F)$. The form B is nondegenerate if and only if this homomorphism is injective. If we represent this homomorphism by a matrix, using the basis $\mathcal{V}$ and the dual basis for $\hom_F(V, F)$, then this matrix is $(B(v_i, v_j))$. Therefore, B is nondegenerate if and only if $\operatorname{disc}(B)_{\mathcal{V}} \neq 0$. This condition is independent of the basis, by the change of basis formula above for the discriminant.

Example 12.18 We now show that the discriminant of a field extension is the discriminant of the trace form. Let K be a finite separable extension of F. Let $B : K \times K \to F$ be defined by $B(a, b) = T_{K/F}(ab)$. Then B is a bilinear form because the trace is linear. The discriminant of B relative to a basis $\mathcal{V} = \{v_1, \ldots, v_n\}$ is $\det(T_{K/F}(v_i v_j))$. But, by Lemma 12.12, this is the discriminant of K/F. Therefore, the previous notions of discriminant are special cases of the notion of discriminant of a bilinear form.

Problems

1. Let $B : V \times V \to F$ be a bilinear form. If $\mathcal{V} = \{v_1, \dots, v_n\}$ is a basis for V, another basis $\mathcal{W} = \{w_1, \dots, w_n\}$ is called a *dual basis to* $\mathcal{V}$ provided that $B(v_i, w_i) = 1$ for all i, and $B(v_i, w_j) = 0$ whenever $i \neq j$. If $\mathcal{V}$ and $\mathcal{W}$ are dual bases, show that $\mathrm{disc}(B)_{\mathcal{V}} \cdot \mathrm{disc}(B)_{\mathcal{W}} = 1$.

2. If B is a nondegenerate bilinear form on V, show that any basis has a dual basis.

3. Let $\{e_i\}$ be a basis for F^n, and choose an $a_i \in F$ for each i. Define B on this basis by $B(e_i, e_j) = 0$ if $i \neq j$ and $B(e_i, e_i) = a_i \in F$. Prove that this function extends uniquely to a bilinear form $B : F^n \times F^n \to F$, and determine the discriminant of B.

4. Let A be a symmetric $n \times n$ matrix, and define a map $B : F^n \times F^n \to F$ by $B(v, w) = vAw^t$, where v and w are viewed as row vectors. Show that B is bilinear. Using the fact that a symmetric matrix can be diagonalized by an orthogonal transformation, use the previous problem to determine the discriminant of B in terms of A.

The remaining problems investigate the use of discriminants in algebraic number theory. They require knowledge of integrality and the Noetherian condition for commutative rings.

5. Let K be a finite separable extension of F, and let A be an integrally closed ring with quotient field F. Let B be the integral closure of A in K. Show that there is an F-basis $v_1, v_2, \dots, v_n$ of K such that $B \subseteq \sum Av_i$.
(Hint: First find a basis $\{w_i\} \subseteq B$, and then use a dual basis relative to the trace form.)

6. Let K be an algebraic number field, and let B be the integral closure of $\mathbb{Z}$ in K. Use the previous problem to show that B is a finitely generated $\mathbb{Z}$-module, and conclude that B is a Noetherian ring. Moreover, show that there is a basis of K that is also a basis for B as a $\mathbb{Z}$-module. Such a basis is called an *integral basis for* $B/\mathbb{Z}$.

7. With the notation of the previous problem, let d be the discriminant of K/F relative to an integral basis $\{v_1, \dots, v_n\}$ of $B/\mathbb{Z}$. Prove that $d \in \mathbb{Z}$. The integer d is called the *discriminant of* $B/\mathbb{Z}$. Show that if we use a different integral basis, then the two discriminants are equal. (One use of discriminants in algebraic number theory is the following: It is known that any nonzero ideal of B factors uniquely into a product of prime ideals. If $P = p\mathbb{Z}$ is a prime ideal of $\mathbb{Z}$, then $PB = Q_1^{e_1} \cdots Q_g^{e_g}$ for some prime ideals Q_i of B and $e_i \geq 1$. Then each $e_i = 1$ if and only if p does not divide d.)

8. Calculate the discriminant of $B/\mathbb{Z}$ for the following fields, where B is the integral closure of $\mathbb{Z}$ in that field.

   (a) $\mathbb{Q}(\sqrt{-1})$.

   (b) $\mathbb{Q}(\sqrt{d})$, where $d > 0$ is a square-free integer.

   (c) $\mathbb{Q}(\omega)$, where ω is a primitive nth root of unity.
   (Hint: Try to prove that $B = \mathbb{Z}[\omega]$. Calculate the discriminant using norms. Show that $N_{K/\mathbb{Q}}(1-\omega) = p$.)

13 Polynomials of Degree 3 and 4

In this section, we show how to determine the Galois group and the roots of an irreducible polynomial of degree 2, 3, or 4. We assume throughout that our polynomials are separable. For degree 2, 3, or 4, requiring that the base field F does not have characteristic 2 or 3 is sufficient to ensure separability. Let $f(x) \in F[x]$ be separable and irreducible over F, and let K be the splitting field over F of f. Set $f(x) = (x - \alpha_1)\cdots(x - \alpha_n) \in K[x]$. If $n = \deg(f)$, note that n divides $[K : F] = |\mathrm{Gal}(K/F)|$, since $[F(\alpha_1) : F] = n$. The Galois group $\mathrm{Gal}(K/F)$ is isomorphic to a subgroup of S_n by identifying S_n as the group of all permutations of the roots of f. Furthermore, $\mathrm{Gal}(K/F)$ is isomorphic to a *transitive subgroup* of S_n; that is, for each pair $i, j \in \{x_1, x_2, \ldots, x_n\}$, there is a $\sigma \in \mathrm{Gal}(K/F)$ with $\sigma(x_i) = x_j$. This fact is due to the isomorphism extension theorem. This limits the possible subgroups of S_n that can be isomorphic to such a Galois group. We call $\mathrm{Gal}(K/F)$ the *Galois group of* f in this section for convenience.

For polynomials of degree 2, there is not much to say. If $f(x) = x^2 + bx + c \in F[x]$ is separable and irreducible over F, then the Galois group of f is S_2, a cyclic group of order 2. If $\mathrm{char}(F) \neq 2$, the quadratic formula can be used to find the roots of f. These roots are $\frac{1}{2}(-b \pm \sqrt{b^2 - 4c})$. Therefore, the splitting field K of f over F is $F(\sqrt{b^2 - 4c})$.

Cubic polynomials

We now consider irreducible polynomials of degree 3. Let f be an irreducible, separable polynomial of degree 3 over a field F, and let K be the splitting field of f over F. Then $\mathrm{Gal}(K/F)$ is isomorphic to a subgroup of S_3. Furthermore, as noted above, $|\,\mathrm{Gal}(K/F)|$ is a multiple of 3. Thus, the only possibilities for $\mathrm{Gal}(K/F)$ are A_3 and S_3. The following theorem is a direct consequence of the results about discriminants in Section 12.

Theorem 13.1 *Let $f(x) \in F[x]$ be an irreducible, separable polynomial of degree 3 over F, and let K be the splitting field of f over F. If D is*

the discriminant of f, then $\mathrm{Gal}(K/F) \cong S_3$ *if and only if* $D \notin F^2$, *and* $\mathrm{Gal}(K/F) \cong A_3$ *if and only if* $D \in F^2$.

Proof. Let $G = \mathrm{Gal}(K/F)$. By Corollary 12.4, $G \subseteq A_3$ if and only if $D \in F^2$. But $G \cong S_3$ or $G \cong A_3$, so $G \cong S_3$ if and only if D is a square in F. □

Example 13.2 The polynomial $x^3 - 3x + 1 \in \mathbb{Q}[x]$ has discriminant $81 = 9^2$, and it is irreducible over $\mathbb{Q}$ by an application of the rational root test. Thus, the Galois group of its splitting field over $\mathbb{Q}$ is A_3. The polynomial $x^3 - 4x + 2$ has discriminant $148 = 2^2 \cdot 37$, so the corresponding Galois group is S_3.

We now present a solution of an arbitrary cubic equation that appeared in Cardano [3] in 1545. We assume that the characteristic of F is neither 2 nor 3. Let $f(x) = x^3 + px + q$. As indicated in Example 12.8, it is sufficient to work with a polynomial of this form, for if $g(x) = x^3 + ax^2 + bx + c$, then by setting $y = x - a/3$, Taylor expansion gives

$$g(x) = g(a/3) + g'(a/3)y + \frac{1}{2}g''(a/3)y^2 + \frac{1}{6}g'''(a/3)y^3,$$

and y is chosen as such because $g''(a/3) = 0$.

Cardano's method is to solve $f = 0$ by writing $x = u + v$ and obtaining two equations in u and v. Replacing x by $u + v$ in the equation $f = 0$ gives

$$u^3 + v^3 + q + (3uv + p)(u + v) = 0.$$

We set $u^3 + v^3 + q = 0$ and $3uv + p = 0$. Thus, $v = -p/(3u)$. Using this in the first equation and multiplying by u^3 yields $4u^6 + qu^3 - p^3/27 = 0$. This is a quadratic equation in u^3, so

$$u^3 = \frac{-q \pm \sqrt{q^2 + 4p^3/27}}{2} = -q/2 \pm \sqrt{\Gamma},$$

where $\Gamma = q^2/4 + p^3/27$. Note that the discriminant D of f is $-4p^3 - 27q^2$, so $\Gamma = -D/108$. Set $A = -q/2 + \sqrt{\Gamma}$ and $B = -q/2 - \sqrt{\Gamma}$. By symmetry of u and v, we may set $u^3 = A$ and $v^3 = B$. Let ω be a primitive third root of unity. The choices for u and v are then

$$u = \sqrt[3]{A},\ \omega\sqrt[3]{A},\ \omega^2\sqrt[3]{A},$$
$$v = \sqrt[3]{B},\ \omega\sqrt[3]{B},\ \omega^2\sqrt[3]{B}.$$

We must choose the cube roots of A and B so that $\sqrt[3]{A}\sqrt[3]{B} = -p/3$. Doing so, the roots of f are

$$\sqrt[3]{A} + \sqrt[3]{B},\ \ \omega\sqrt[3]{A} + \omega^2\sqrt[3]{B},\ \ \omega^2\sqrt[3]{A} + \omega\sqrt[3]{B}.$$

Example 13.3 Consider $x^3 - 3x + 1$. Then $\Gamma = -D/108 = -81/108 = -3/4$. We have $p = -3$ and $q = 1$. Then $A = -1/2 + i\sqrt{3}/2$ and $B = -1/2 - i\sqrt{3}/2$, so $A = \exp(2\pi i/3)$ and $B = \exp(-2\pi i/3)$. We can then set $u = \exp(2\pi i/9)$ and $v = \exp(-2\pi i/9)$. Also, $\omega = \exp(2\pi i/3)$. By simplifying the formulas for the roots of f, we see that the three roots are $2\cos(2\pi/9), 2\cos(8\pi/9)$, and $2\cos(14\pi/9)$.

Suppose that the polynomial $f(x) = x^3 + px + q$ has real coefficients. If $\Gamma > 0$, then $D < 0$, so D is not a square in F. We can then take the real cube roots of A and B for u and v. Furthermore, if $\omega = (-1 + i\sqrt{3})/2$, we see that the three roots of f are

$$\alpha_1 = \sqrt[3]{A} + \sqrt[3]{B} \in \mathbb{R},$$

$$\alpha_2 = -\left(\frac{\sqrt[3]{A} + \sqrt[3]{B}}{2}\right) + i\sqrt{3}\left(\frac{\sqrt[3]{A} - \sqrt[3]{B}}{2}\right),$$

and

$$\alpha_3 = -\left(\frac{\sqrt[3]{A} + \sqrt[3]{B}}{2}\right) - i\sqrt{3}\left(\frac{\sqrt[3]{A} - \sqrt[3]{B}}{2}\right).$$

On the other hand, if $\Gamma < 0$, then $A = -q/2 + i\sqrt{-\Gamma}$ and $B = -q/2 - i\sqrt{-\Gamma}$. If we choose $\sqrt[3]{A} = a + bi$ to satisfy $\sqrt[3]{A}\sqrt[3]{B} = -p/3$, we must then have $\sqrt[3]{B} = a - bi$. The roots of f are then $\alpha_1 = 2a$, $\alpha_2 = -a - b\sqrt{3}$, and $\alpha_3 = -a + b\sqrt{3}$, and all three are real numbers.

The case where $\Gamma < 0$ historically had been called the "irreducible case," since it was realized that even though all three roots are real, the roots cannot be expressed in terms of real radicals.

Quartic polynomials

We now consider polynomials of degree 4. Let $f(x) = x^4 + ax^3 + bx^2 + cx + d$ be an irreducible, separable polynomial over a field F, and let f factor as

$$f(x) = (x - \alpha_1)(x - \alpha_2)(x - \alpha_3)(x - \alpha_4)$$

in some splitting field. The key idea we use to find the roots and the Galois group G of f is to work with an associated cubic polynomial. Set

$$\beta_1 = \alpha_1\alpha_2 + \alpha_3\alpha_4,$$
$$\beta_2 = \alpha_1\alpha_3 + \alpha_2\alpha_4,$$
$$\beta_3 = \alpha_1\alpha_4 + \alpha_2\alpha_3,$$

and

$$r(x) = (x - \beta_1)(x - \beta_2)(x - \beta_3).$$

A computation shows that

$$r(x) = x^3 - bx^2 + (ac - 4d)x + 4bd - a^2d - c^2 \in F[x].$$

The polynomial r is called the *resolvent* of f. An easy calculation shows that f and r have the same discriminant. Let $K = F(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$, a splitting field of f over F, and let $L = F(\beta_1, \beta_2, \beta_3)$, a splitting field of r over F. Note that L/F is Galois. Let

$$V = \{e, (12)(34), (13)(24), (14)(23)\},$$

a subgroup of S_4 of order 4. Then $V \subseteq A_4$ and V is normal in S_4. Each β_i is fixed by V, so $L \subseteq \mathcal{F}(G \cap V)$. The reverse inclusion is also true, which can be seen by showing that any element of $G - G \cap V$ moves one of the β_i. The group G is isomorphic to a transitive subgroup of S_4, and it has order a multiple of 4. It is not hard to show that the transitive subgroups of S_4 of order 24 and 12, respectively, are S_4 and A_4, and that the transitive subgroups of order 4 are V and the cyclic subgroups generated by a 4-cycle. The subgroup generated by (1234) and (24) is a transitive subgroup of order 8. Since this is a 2-Sylow subgroup of S_4, any subgroup of order 8 is isomorphic to it, and so is isomorphic to D_4, the *dihedral group* of order 8. We write C_4 for the unique up to isomorphism cyclic group of order 4. We now show how to determine G in terms of the discriminant of f and the resolvent r. The particular statement of the following theorem we give appeared in Kappe and Warren [18].

Theorem 13.4 *With the notation above, let $m = [L : F]$.*

1. *$G \cong S_4$ if and only if $r(x)$ is irreducible over F and $D \notin F^2$, if and only if $m = 6$.*

2. *$G \cong A_4$ if and only if $r(x)$ is irreducible over F and $D \in F^2$, if and only if $m = 3$.*

3. *$G \cong V$ if and only if $r(x)$ splits over F, if and only if $m = 1$.*

4. *$G \cong C_4$ if and only if $r(x)$ has a unique root $t \in F$ and $h(x) = (x^2 - tx + d)(x^2 + ax + (b - t))$ splits over L, if and only if $m = 2$ and $f(x)$ is reducible over L.*

5. *$G \cong D_4$ if and only if $r(x)$ has a unique root $t \in F$ and $h(x)$ does not split over L, if and only if $m = 2$ and f is irreducible over L.*

Proof. We first point out a couple of things. First, $[K : L] \leq 4$, since $K = L(\alpha_1)$. This equality follows from the fundamental theorem, since only the identity automorphism fixes $L(\alpha_1)$. Second, $r(x)$ is irreducible over F if and only if $m = 3$ or $m = 6$. Also, $r(x)$ has a unique root in F if and only if $m = 2$. Finally, if σ is a 4-cycle, then $\sigma^2 \in V$.

Suppose that $r(x)$ is irreducible over F. Then m is either 3 or 6, so 3 divides $|G|$. This forces G to be isomorphic to either S_4 or A_4. In either case, $V \subseteq G$, so $L = \mathcal{F}(V)$ by the fundamental theorem. Thus, $[K : L] = 4$,

so $G = S_4$ if $m = 6$, and $G = A_4$ if $m = 3$. Alternatively, $G = S_4$ if and only if $D \notin F^2$, and $G = A_4$ if and only if $D \in F^2$. Conversely, if $G = S_4$, then $m = |S_4 : V| = 6$, and if $G = A_4$, then $m = |A_4 : V| = 3$. In either case, 3 divides $|G|$, so $r(x)$ is irreducible over F.

Next, $r(x)$ splits over F if and only if $L = F$, if and only if $m = 1$. If this occurs, then L corresponds to both G and $G \cap V$, so $G \subseteq V$. Since $|G|$ is a multiple of 4, we see $G = V$. Conversely, if $G = V$, then L corresponds to $G \cap V = G$, so $L = F$; thus, $m = 1$ and $r(x)$ splits over F.

For the final case, we suppose that $r(x)$ has a single root t in F. This is equivalent to $m = 2$. Thus, $|G : G \cap V| = 2$, so $G \not\subseteq V$. The only possibilities for G are $G \cong C_4$ or $G \cong D_4$. Conversely, if G is either isomorphic to D_4 or C_4, then $m = |G : G \cap V| = 2$, so $r(x)$ has a unique root F. Now f is irreducible over L if and only if $[K : L] = 4$, if and only if $[K : F] = 8$, if and only if $G \cong D_4$. Therefore, $G \cong C_4$ if and only if f is reducible over L. By relabeling if necessary, we may suppose that $t = \alpha_1\alpha_2 + \alpha_3\alpha_4$. Then $h(x)$ factors over K as

$$h(x) = (x - \alpha_1\alpha_2)(x - \alpha_3\alpha_4)(x - (\alpha_1 + \alpha_2))(x - (\alpha_3 + \alpha_4)).$$

If h splits over L, then $\alpha_1 + \alpha_2$ and $\alpha_1\alpha_2$ are in L. Thus, α_1 satisfies the quadratic polynomial

$$x^2 - (\alpha_1 + \alpha_2)x + \alpha_1\alpha_2 = (x - \alpha_1)(x - \alpha_2) \in L[x].$$

Thus, $[K : L] \leq 2$ because $K = L(\alpha_1)$. Therefore, $[K : F] \leq 4$, so $G \cong C_4$. If $G \cong C_4$, let σ be a generator for G. Then $\sigma^2 \in G \cap V$, since L is the unique nontrivial subfield of K/F. To fix $t = \alpha_1\alpha_2 + \alpha_3\alpha_4$, we must have $\sigma^2 = (12)(34)$. Then $\alpha_1 + \alpha_2$, $\alpha_3 + \alpha_4$, $\alpha_1\alpha_2$, and $\alpha_3\alpha_4$ are all fixed by σ^2, so they lie in L. Thus, h splits over L. This completes the proof of the theorem. □

We now find the roots of the general polynomial of degree 4. We point out that the formulas we derive below only require us to find one root of the resolvent polynomial, and such a root can be found by Cardano's method. Our approach is not that of Ferrari, a student of Cardano and the first to solve the quartic, although deep down it is much the same. His method is addressed in Problem 1. Instead, our method is based on the theorem of Galois, which says that there is an algebraic formula for the roots of a polynomial if and only if the Galois group of the polynomial is a solvable group. We shall discuss this theorem in detail in Section 16. To use hindsight, the idea is that given a sequence of subgroups $G \supseteq H_1 \supseteq \cdots \supseteq H_t = \langle \mathrm{id} \rangle$ for which H_{i+1} is normal in H_i with H_i/H_{i+1} Abelian, which exists for a solvable group, we obtain a sequence of intermediate subfields $F = L_t \subseteq L_{t-1} \subseteq \cdots \subseteq K$ for which the extension L_{i-1}/L_i is easy to describe. By describing L_{t-1}, then L_{t-2}, and so on, eventually we describe K. This brings up the question of how to motivate the definition

of the resolvent polynomial. For S_4, a natural chain of subgroups is $S_4 \supseteq A_4 \supseteq V \supseteq \langle \mathrm{id} \rangle$, since this is the usual sequence that shows S_4 is solvable. If $f(x) = (x - t_1)(x - t_2)(x - t_3)(x - t_4)$, then the automorphisms in V fix $t_1t_2 + t_3t_4$, $t_1t_3 + t_2t_4$, and $t_1t_4 + t_2t_3$, and we have seen that the fixed field of V is the field generated by these three elements. This field is then the splitting field of the polynomial whose three roots are these three elements; that is, it is the splitting field of the resolvent of f.

Let us now find the roots of the general fourth degree polynomial. Let k be a field of characteristic not 2, and let $K = k(t_1, t_2, t_3, t_4)$ be the rational function field in four variables over k. Let

$$\begin{aligned} f(x) &= (x - t_1)(x - t_2)(x - t_3)(x - t_4) \\ &= x^4 + ax^3 + bx^2 + cx + d \in k(s_1, s_2, s_3, s_4)[x], \end{aligned}$$

where s_i is the ith elementary symmetric polynomial in the t_j. Then $s_1 = -a$, $s_2 = b$, $s_3 = -c$, and $s_4 = d$. Recall from Example 3.9 that if $F = k(s_1, s_2, s_3, s_4)$, then $K = F(t_1, t_2, t_3, t_4)$ is the splitting field over F of f, and $\mathrm{Gal}(K/F) = S_4$. Set

$$\begin{aligned} \beta_1 &= t_1t_2 + t_3t_4, \\ \beta_2 &= t_1t_3 + t_2t_4, \\ \beta_3 &= t_1t_4 + t_2t_3. \end{aligned}$$

The resolvent r is

$$\begin{aligned} r(x) &= (x - \beta_1)(x - \beta_2)(x - \beta_3) \\ &= x^3 - bx^2 + (ac - 4d)x + 4bd - a^2d - c^2. \end{aligned}$$

Let $L = F(\beta_1, \beta_2, \beta_3)$, the fixed field of V. For simplicity, we write $\sigma_1 = (12)(34)$, $\sigma_2 = (13)(24)$, and $\sigma_3 = (14)(23)$. Let $u = (t_1 + t_2) - (t_3 + t_4)$. Then $\sigma_1(u) = u$ and $\sigma_i(u) = -u$ for $i = 2, 3$. Therefore, $u^2 \in L$. Let $M = L(u)$. Then M corresponds to $\{\mathrm{id}, \sigma_1\}$. Finally, let $v = t_1 - t_2$. Then $\sigma_1(v) = -v$, so $v^2 \in M$. Also, $M(v)$ is fixed only by id, so $K = M(v)$. We have

$$\begin{aligned} u^2 &= (t_1 + t_2)^2 + (t_3 + t_4)^2 - 2(t_1 + t_2)(t_3 + t_4) \\ &= t_1^2 + t_2^2 + t_3^2 + t_4^2 + 2(t_1t_2 + t_3t_4) - 2(t_1t_3 + t_2t_4 + t_1t_4 + t_2t_3) \\ &= s_1^2 - 2s_2 + 2\beta_1 - 2(\beta_2 + \beta_3) = s_1^2 - 2s_2 + 4\beta_1 - 2b \\ &= a^2 - 4b + 4\beta_1. \end{aligned}$$

To determine v^2, we first point out that $u + s_1 = 2(t_1 + t_2)$, so $t_1 + t_2 = \frac{1}{2}(s_1 + u)$. Similarly, $t_3 + t_4 = \frac{1}{2}(s_1 - u)$. Now,

$$\begin{aligned} v^2 &= (t_1 - t_2)^2 = (t_1 + t_2)^2 - 4t_1t_2 = \frac{1}{4}(s_1 + u)^2 - 4t_1t_2 \\ &= \frac{1}{4}(-a + u)^2 - 4t_1t_2. \end{aligned}$$

However, we can determine t_1t_2 in terms of the coefficients as follows. If we expand $(t_1t_2 - t_3t_4)u$, recalling that $u = (t_1 + t_2) - (t_3 + t_4)$, we get

$$\begin{aligned}
&(t_1t_2 - t_3t_4)((t_1 + t_2) - (t_3 + t_4)) \\
&= t_1^2t_2 + t_1t_2^2 + t_3^2t_4 + t_3t_4^2 - (t_1t_2t_3 + t_1t_2t_4 + t_2t_3t_4 + t_1t_3t_4) \\
&= (t_1t_2 + t_3t_4)(t_1 + t_2 + t_3 + t_4) - 2s_3 \\
&= s_1\beta_1 - 2s_3 = -a\beta_1 + 2c.
\end{aligned}$$

Thus, $t_1t_2 - t_3t_4 = u^{-1}(2c - a\beta_1)$. Since $\beta_1 = t_1t_2 + t_3t_4$, we see that

$$\begin{aligned}
t_1t_2 &= \frac{1}{2}\left(\beta_1 + \frac{1}{u}(2c - a\beta_1)\right), \\
t_3t_4 &= \frac{1}{2}\left(\beta_1 - \frac{1}{u}(2c - a\beta_1)\right),
\end{aligned}$$

so

$$v^2 = \frac{1}{4}(u-a)^2 - 2\left(\beta_1 + \frac{1}{u}(2c - a\beta_1)\right).$$

Once we have a formula for t_1, we will have formulas for the other t_i, since $t_2 = \sigma_1(t_1)$, $t_3 = \sigma_2(t_1)$, and $t_4 = \sigma_3(t_1)$. To find t_1, note that

$$t_1 = \frac{1}{2}(t_1 + t_2 + t_1 - t_2) = \frac{1}{2}\left(v + \frac{1}{2}(u-a)\right).$$

To get formulas for t_2, t_3, and t_4, we need to know $\sigma_i(v)$. We have $\sigma_1(v) = -v$. Let

$$v' = t_3 - t_4 = \sigma_2(v) = \sigma_3(v).$$

Since $\sigma_1(u) = u$, $\sigma_2(u) = -u$, and $\sigma_3(u) = -u$, we see that

$$(v')^2 = \frac{1}{4}(u-a)^2 - 2\left(\beta_1 - \frac{1}{u}(2c - a\beta_1)\right).$$

Therefore, we have

$$\begin{aligned}
t_1 &= \frac{1}{2}\left(v + \frac{1}{2}(u-a)\right), \\
t_2 &= \frac{1}{2}\left(-v + \frac{1}{2}(u-a)\right), \\
t_3 &= \frac{1}{2}\left(v' + \frac{1}{2}(-u-a)\right), \\
t_4 &= \frac{1}{2}\left(-v' + \frac{1}{2}(-u-a)\right).
\end{aligned}$$

For a specific polynomial, these formulas will work provided that $u \neq 0$. Since the roots of $r(x)$ are distinct, provided that f has no repeated roots, at most one choice of β will make $u = 0$.

Recall that the Galois group of f over F is S_4. Figures 13.1 and 13.2 list some of the subgroups of S_4 and the corresponding intermediate subfields. To make the diagrams manageable, we list only one subgroup/subfield of each "type." For instance, there are three subgroups generated by a 4-cycle, and six subgroups generated by a 3-cycle. We list only one of each. The group $S(1)$ below is the group of permutations that fix 1, and the element Δ is the element $\prod_{i<j}(t_i - t_j)$, so Δ^2 is the discriminant of f and also of r.

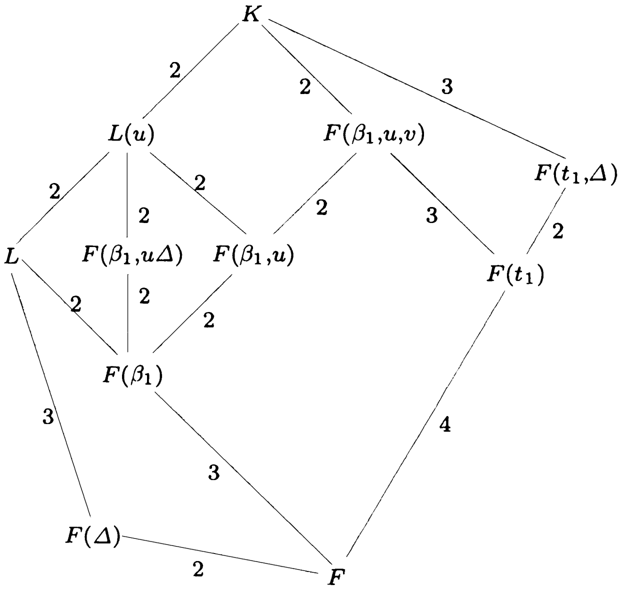


FIGURE 13.1. Field tower for $F(t_1, t_2, t_3, t_4)/F$.

Example 13.5 Let $f(x) = x^4 + x^3 + x^2 + x + 1$. Then $a = b = c = d = 1$, so $s_1 = s_3 = -1$ and $s_2 = s_4 = 1$. Also,

$$r(x) = x_3 - x_2 - 3x + 2 = (x - 2)(x^2 + x - 1).$$

Set $\beta_1 = 2$. Then $u = \sqrt{5}$. Also,

$$\begin{aligned} v^2 &= \frac{1}{4}(-1 + u)^2 - 2(2 + u^{-1}(-2 + 2)) \\ &= \frac{1}{4}(u^2 - 2u + 1) - 4 = -\frac{5 + u}{2}. \end{aligned}$$

Thus, $v = \frac{i}{2}\sqrt{10 - 2\sqrt{5}}$. In addition, we see that $v' = \frac{i}{2}\sqrt{10 - 2\sqrt{5}}$. The roots of f are then

$$\frac{1}{2}\left(\frac{i}{2}\sqrt{10 + 2\sqrt{5}} + \frac{1}{2}(-1 + \sqrt{5})\right) = \frac{1}{4}(-1 + \sqrt{5}) + \frac{i}{4}\sqrt{10 + 2\sqrt{5}},$$

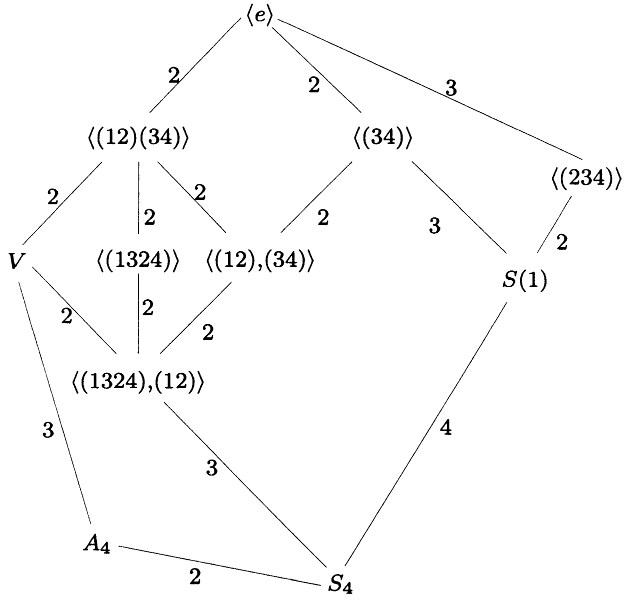


FIGURE 13.2. Group tower for S_4.

$$\frac{1}{2}\left(-\frac{i}{2}\sqrt{10+2\sqrt{5}}+\frac{1}{2}(-1+\sqrt{5})\right)=\frac{1}{4}(-1+\sqrt{5})-\frac{i}{4}\sqrt{10+2\sqrt{5}},$$
$$\frac{1}{2}\left(\frac{i}{2}\sqrt{10-2\sqrt{5}}+\frac{1}{2}(-1-\sqrt{5})\right)=\frac{1}{4}(-1-\sqrt{5})+\frac{i}{4}\sqrt{10-2\sqrt{5}},$$
$$\frac{1}{2}\left(-\frac{i}{2}\sqrt{10-2\sqrt{5}}+\frac{1}{2}(-1-\sqrt{5})\right)=\frac{1}{4}(-1-\sqrt{5})-\frac{i}{4}\sqrt{10-2\sqrt{5}}.$$

The polynomial $h(x)=(x^2-2x+1)(x^2+x-1)$ splits over L, so by Theorem 13.4 the Galois group of f is isomorphic to C_4. Alternatively, $f(x)$ is the fifth cyclotomic polynomial $\Psi_5(x)$, so Section 7 tells us that the Galois group of f is cyclic.

Example 13.6 Let $f(x)=x^4-4x^3+4x^2+6$. This polynomial is irreducible by the Eisenstein criterion. Now,

$$r(x)=x^3-4x^2-24x=x(x^2-4x-24),$$

so $L=\mathbb{Q}(\sqrt{7})$. Take $\beta_1=0$. Then

$$h(x)=(x^2+6)(x^2-4x+4)=(x^2+6)(x-2)^2.$$

Since h does not split over L, we see that the Galois group of f is isomorphic to D_4.

Example 13.7 Let p be a prime, and let $f(x) = x^4 + px + p$. Then $r(x) = x^3 - 4px - p^2$. To test for roots of $r(x)$ in $\mathbb{Q}$, we only need to check $\pm 1, \pm p, \pm p^2$. We see that ± 1 and $\pm p^2$ are never roots, but $r(p) = p^2(p-5)$ and $r(-p) = p^2(3-p)$. Therefore, for $p \neq 3, 5$, the resolvent r has no roots in $\mathbb{Q}$; hence, r is irreducible over $\mathbb{Q}$. The discriminant $D = p^3(256 - 27p)$ is not a square in $\mathbb{Q}$, since if p is odd, then p does not divide $256 - 27p$, and $D = 1616 \notin \mathbb{Q}^2$ for $p = 2$. Let G be the Galois group of f. Then $G \cong S_4$ for $p \neq 3, 5$. If $p = 3$, let $\beta_1 = -3$. Then $r(x) = (x+3)(x^2 - 3x - 3)$, so $L = \mathbb{Q}(\sqrt{21})$. Then $h(x) = (x^2 + 3x + 3)(x^2 + 3)$ does not split over L, so $G \cong D_4$. If $p = 5$, then $r(x) = (x-5)(x^2 + 5x + 5)$, so $L = \mathbb{Q}(\sqrt{5})$. As $h(x) = (x^2 - 5x + 5)(x^2 - 5)$, h splits over L, so $G \cong C_4$.

Example 13.8 Let $l \in \mathbb{Q}$, and let $f(x) = x^4 - l$. Then the resolvent of f is $r(x) = x^3 + 4lx = x(x^2 + 4l)$. If $-l$ is not a square in $\mathbb{Q}$, then $r(x)$ has exactly one root in $\mathbb{Q}$. Moreover, $h(x) = x^2(x^2 + l)$ does not factor completely over $\mathbb{Q}$, so the Galois group G of f is D_4 by Theorem 13.4. On the other hand, if $-l$ is a square in $\mathbb{Q}$, then r factors completely over $\mathbb{Q}$, so $G \cong V$. For example, the Galois group of $x^4 + 4$ is V. The splitting field of $x^4 + 4$ over $\mathbb{Q}$ is then $\mathbb{Q}(\sqrt[4]{-4})$.

Problems

1. *Ferrari's solution of the quartic.* Here is Ferrari's method for finding the roots of a quartic, which appeared in [3]. Let $g(x) = x^4 + ax^3 + bx^2 + cx + d$. Starting with $g(x) = 0$, move the quadratic part of f to the right-hand side. Show by completion of squares that the equation becomes

$$\left(x^2 + \frac{1}{2}ax\right)^2 = \left(\frac{1}{4}a^2 - b\right)x^2 - cx - d.$$

Ferrari's idea is to add to both sides the expression $y(x^2 + ax/2) + y^2/4$ for some y, so that the left-hand side is a perfect square. The equation then becomes

$$\left(x^2 + \frac{1}{2}ax + \frac{1}{2}y\right)^2 = \left(\frac{1}{4}a^2 - b + y\right)x^2 + \left(\frac{1}{2}ay - c\right)x + \frac{1}{4}y^2 - d.$$

We wish to choose y so that the right-hand side becomes a square, $(ex+f)^2$. Writing the right-hand side as $Ax^2 + Bx + C$, this is possible if and only if $B^2 - 4AC = 0$. Show that this gives an equation in y to be solved, and if r is the resolvent of g, then this equation is $r(x) = 0$. Given such a y, take the equation

$$\left(x^2 + \frac{1}{2}ax + \frac{1}{2}y\right)^2 = (ex + f)^2 \qquad (13.1)$$

and obtain two quadratic equations in x and solve them to find the general solution to $g(x) = 0$. Relate Ferrari's method by the method of the section by showing that $e = \frac{1}{2}u$ and that the discriminants of the two quadratic equations in (13.1) are equal to v^2 and $(v')^2$.

2. Solve $x^4 + 4x - 1 = 0$ by Ferrari's method and by the method of the section.

3. Show that $2\cos(2\pi/15)$ is a root of $x^4 - x^3 - 4x^2 + 4x + 1$. What are the other roots?

4. Solve the equation $\left((x+2)^2 + x^2\right)^3 = 8x^4(x+2)^2$ by setting $y = x+1$.

5. Find the roots of $x^4 + px^3 + qx^2 + px + 1$, and notice that cube and fourth roots are not needed.

6. Use the ideas of this section to show that $\sqrt[3]{\sqrt{5}+2} - \sqrt[3]{\sqrt{5}-2} = 1$ and that $\sqrt[3]{7+\sqrt{50}} + \sqrt[3]{7-\sqrt{50}} = 2$.

7. Find the roots of $x^3 - 6x - 6$ and the roots of $2x^3 + 6x + 3$.

8. If the specific gravity of cork is 0.25, to what depth will a sphere of radius r made of cork sink in water? Archimedes' principle is that the weight of water displaced is equal to the weight of the cork. (You might want to ask yourself why this problem is here!)

9. Let $f(x) = x^4 + ax^2 + b \in \mathbb{Q}[x]$. Determine the Galois group of f.

10. Let K be a field extension of F with $[K : F] = 4$. Show that K contains an intermediate subfield L with $[L : F] = 2$ if and only if $K = F(\alpha)$, where α satisfies a polynomial $x^4 + ax^2 + b \in F[x]$.

11. Given the splitting field $k(t_1, t_2, t_3, t_4)$ of the general quartic $(x - t_1)(x - t_2)(x - t_3)(x - t_4)$ over $k(s_1, s_2, s_3, s_4)$, for each pair L_2/L_1 of intermediate subfields for which there is no proper intermediate subfield, find a single element that generates L_2 over L_1, and find this element's minimal polynomial over L_1.

14 The Transcendence of π and e

The two best known and most important nonrational real numbers are π and e. In this section, we will show that both of these numbers are transcendental over $\mathbb{Q}$. In Section 15, we will use the transcendence of π to prove that it is impossible to square the circle, one of the ruler and compass construction questions of ancient Greece that remained unsolved for 2500 years.

The recognition that irrational numbers exist can be traced back to the Pythagoreans' proof over 2000 years ago that $\sqrt{2}$ is irrational. However, it was not known whether π was rational until 1761, when Lambert proved that π is irrational. Euler, after finding a continued fraction expression for e, believed that e was irrational but was not able to prove it. In 1767, Lambert gave a proof that e was irrational. By this time, people suspected that not all numbers were algebraic. The existence of transcendental numbers remained an open question until Liouville in 1844 came up with a criterion for a complex number to be algebraic and showed that transcendental numbers do exist. Liouville's method showed that numbers whose decimal expansion contained increasingly long strings of 0's are transcendental. For instance, his method showed that $\sum_{n=0}^{\infty} 10^{-n!}$ is transcendental. Proving that a particular number, such as π and e, is transcendental is another matter. The transcendence of e was not proved until 1873, when Hermite gave a proof. Nine years later, Lindemann used Hermite's method to prove that π is transcendental.

In this section, we give a more general result of Lindemann that implies the transcendence of both e and π. A more detailed proof of this result was given by Weierstrauss in 1895 and often goes under the name of the Lindemann–Weierstrauss theorem. Actually, we give an alternative version of this theorem that is a little easier to prove than the original version. The original version is mentioned in Problem 1. The proof of the Lindemann–Weierstrauss theorem requires some analysis, including complex integration, along with Galois theory.

Theorem 14.1 (Lindemann–Weierstrauss) *Let $\alpha_1, \dots, \alpha_m$ be distinct algebraic numbers. Then the exponentials $e^{\alpha_1}, \dots, e^{\alpha_m}$ are linearly independent over $\mathbb{Q}$.*

Corollary 14.2 *The numbers π and e are transcendental over $\mathbb{Q}$.*

Proof of the corollary. Suppose that e is algebraic over $\mathbb{Q}$. Then there are rationals r_i with $\sum_{i=0}^{n} r_i e^i = 0$. This means that the numbers e^0, $e^1, \dots, e^{n-1}$ are linearly dependent over $\mathbb{Q}$. By choosing $m = n + 1$ and $\alpha_i = i-1$, this dependence is false by the theorem. Thus, e is transcendental over $\mathbb{Q}$. For π, we note that if π is algebraic over $\mathbb{Q}$, then so is πi; hence, $e^0, e^{\pi i}$ are linearly independent over $\mathbb{Q}$, which is false since $e^{\pi i} = -1$. Thus, π is transcendental over $\mathbb{Q}$. □

Proof of the theorem. Suppose that there are $a_j \in \mathbb{Q}$ with

$$\sum_{j=1}^{m} a_j e^{\alpha_j} = 0.$$

By multiplying by a suitable integer, we may assume that each $a_j \in \mathbb{Z}$. Moreover, by eliminating terms if necessary, we may also assume that each

$a_j \neq 0$. Let K be the normal closure of $\mathbb{Q}(\alpha_1, \ldots, \alpha_m)/\mathbb{Q}$. Then K is a Galois extension of $\mathbb{Q}$. Suppose that $\mathrm{Gal}(K/\mathbb{Q}) = \{\sigma_1, \ldots, \sigma_n\}$. Since $\sum_{j=1}^{m} a_j e^{\alpha_j} = 0$, we have

$$0 = \prod_{k=1}^{n} \left(\sum_{j=1}^{n} a_j e^{\sigma_k(\alpha_j)} \right) = \sum_{j=0}^{r} c_j e^{\beta_j},$$

where the $c_j \in \mathbb{Z}$ and the β_j can be chosen to be distinct elements of K by gathering together terms with the same exponent. Moreover, some $c_j \neq 0$ (see Problem 4); without loss of generality, say $c_0 \neq 0$. If $\sigma \in \mathrm{Gal}(K/\mathbb{Q})$, then the n terms $\sum_{j=1}^{n} a_j e^{\sigma\sigma_k(\alpha_j)}$ for $1 \leq k \leq n$ are the terms $\sum_{j=1}^{n} a_j e^{\sigma_k(\alpha_j)}$ in some order, so the product is unchanged when replacing $\sigma_k(\alpha_j)$ by $\sigma\sigma_k(\alpha_j)$. Since each β_j is a sum of terms of the form $\sigma_k(\alpha_l)$, the exponents in the expansion of $\prod_{k=1}^{n} \left(\sum_{j=1}^{n} a_j e^{\sigma\sigma_k(\alpha_j)} \right)$ are the various $\sigma(\beta_j)$. Thus, we obtain equations

$$0 = \sum_{j=0}^{r} c_j e^{\sigma_i(\beta_j)}$$

for each i. Multiplying the ith equation by $e^{\sigma_i(\beta_0)}$, we get

$$0 = c_0 + \sum_{j=1}^{r} c_j e^{\sigma_i(\gamma_j)}, \tag{14.1}$$

where $\gamma_j = \beta_j - \beta_0$. Note that $\gamma_j \neq 0$ since the β_j are all distinct. Each $\gamma_j \in K$; hence, each γ_j is algebraic over $\mathbb{Q}$. Thus, for a fixed j, the elements $\sigma_i(\gamma_j)$ are roots of a polynomial $g_j(x) \in \mathbb{Q}[x]$, where the leading coefficient b_j of $g_j(x)$ can be taken to be a positive integer. Moreover, we may assume that $g_j(0) \neq 0$ by using an appropriate multiple of $\min(\mathbb{Q}, \gamma_j)$ for $g_j(x)$.

We now make estimates of some complex integrals. If $f(x)$ is a polynomial, let

$$F(x) = \sum_{i=0}^{\infty} f^{(i)}(x),$$

where $f^{(i)}(x)$ is the ith derivative of f. This sum is finite since f is a polynomial, so F is also a polynomial. Note that $F(x) - F'(x) = f(x)$, so

$$\frac{d}{dx}\left(e^{-x}F(x)\right) = -e^{-x}f(x).$$

Therefore,

$$\int_0^a e^{-x} f(x) dx = F(0) - e^{-a}F(a)$$

or

$$F(a) - e^a F(0) = -e^a \int_0^a e^{-x} f(x)dx.$$

By setting $a = \sigma_i(\gamma_j)$, multiplying by c_j, and summing over i and j, we get

$$\sum_{j=1}^{r}\sum_{i=1}^{n} c_j F(\sigma_i(\gamma_j)) - F(0)\sum_{j=1}^{r}\sum_{i=1}^{n} c_j e^{\sigma_i(\gamma_j)} = -\sum_{j=1}^{r}\sum_{i=1}^{n} c_j e^{\sigma_i(\gamma_j)} \int_0^{\sigma_i(\gamma_j)} e^{-z} f(z)dz.$$

Using Equation (14.1) and rearranging the second sum gives us an equation

$$nc_0F(0) + \sum_{j=1}^{r} c_j \sum_{i=1}^{n} F(\sigma_i(\gamma_j)) = -\sum_{j=1}^{r}\sum_{i=1}^{n} c_j e^{\sigma_i(\gamma_j)} \int_0^{\sigma_i(\gamma_j)} e^{-z} f(z)dz. \tag{14.2}$$

We define f by

$$f(x) = \frac{(b_1 \cdots b_r)^{prn}}{(p-1)!} x^{p-1} \left(\prod_{j=1}^{r} g_j(x)\right)^p,$$

where p is a prime yet to be specified. Recall that b_i is the leading coefficient of $g_i(x)$ and that each b_i is a positive integer. From this definition, we see that

$$0 = f(0) = f'(0) = \cdots = f^{(p-2)}(0)$$

while $f^{(p-1)}(0) = (b_1 \cdots b_r)^{pnr} \prod_{j=1}^{r} g_j(0)^p \neq 0$. We choose p to be any prime larger than $\max_j \{b_j, g_j(0)\}$, so that p does not divide $f^{(p-1)}(0)$. However, for $t \geq p$, the polynomial $f^{(t)}(x)$ can be written in the form

$$f^{(t)}(x) = p(b_1 \cdots b_r)^{prn} h_t(x),$$

where $h_j(x) \in \mathbb{Z}[x]$ has degree at most $prn - 1$. Thus, $f^{(t)}(0)$ is divisible by p for $t \geq p$; hence, $F(0) = f^{(p-1)}(0) + \sum_{j \neq p-1} f^{(j)}(0)$ is not divisible by p. If we further restrict p so that $p > n$ and $p > c_0$, then p does not divide $nc_0F(0)$. We will complete the proof by showing that the first sum in Equation (14.2) is an integer divisible by p and that the right-hand side of Equation (14.2) goes to 0 as p gets large. This will show that the left-hand side is at least 1 in absolute value, which will then give a contradiction.

We now show that $\sum_{j=1}^{r} c_j \sum_{i=1}^{n} F(\sigma_i(\gamma_j))$ is an integer divisible by p. We do this by showing that each term $\sum_{i=1}^{n} F(\sigma_i(\gamma_j))$ is an integer divisible by p. Now,

$$\sum_{i=1}^{n} F(\sigma_i(\gamma_j)) = \sum_{k}\sum_{i=1}^{n} f^{(k)}(\sigma_i(\gamma_j)).$$

Since $g_j(x)^p$ divides $f(x)$ and each $\sigma_i(\gamma_j)$ is a root of $g_j(x)$, we see that

$$0 = f(\sigma_i(\gamma_j)) = f'(\sigma_i(\gamma_j)) = \cdots = f^{(p-1)}(\sigma_i(\gamma_j)).$$

For $t \geq p$, since $f^{(t)}(x) = p(b_1 \cdots b_r)^{pnr} h_j(x)$,

$$\sum_{i=1}^{n} f^{(t)}(\sigma_i(\gamma_j)) = p \cdot \sum_{i=1}^{n} (b_1 \cdots b_r)^{pnr} h_t(\sigma_i(\gamma_j)). \tag{14.3}$$

However, this sum is invariant under the action of $\mathrm{Gal}(K/\mathbb{Q})$, so it is a rational number. Moreover, $\sum_{i=1}^{n}(b_1 \cdots b_r)^{pnr} h_t(x_i)$ is a symmetric polynomial in $x_1, \ldots, x_n$ of degree at most $prn - 1$. The $\sigma_i(\gamma_j)$ are roots of the polynomial $g_j(x)$, whose leading coefficient is b_j, so the second sum in Equation (14.3) is actually an integer by an application of the symmetric function theorem (see Problem 5). This shows that $\sum_{j=1}^{r} c_j \sum_{i=1}^{n} F(\sigma_i(\gamma_j))$ is an integer divisible by p; hence, the left-hand side of Equation (14.2) is a nonzero integer. This means that

$$\left| \sum_{j=1}^{r} \sum_{i=1}^{n} c_j e^{\sigma_i(\gamma_j)} \int_0^{\sigma_i(\gamma_j)} e^{-z} f(z) dz \right| \geq 1.$$

Let

$$\begin{aligned} m_1 &= \max_j \{|c_j|\}, \\ m_2 &= \max_{i,j} \left\{ \left| e^{\sigma_i(\gamma_j)} \right| \right\}, \\ m_3 &= \max_{i,j} \{|\sigma_i(\gamma_j)|\}, \end{aligned}$$

and

$$\begin{aligned} m_4 &= \max_{s \in [0,1]} \left\{ \left| e^{-z} \right| : z = s\sigma_i(\gamma_j) \right\}, \\ m_5 &= \max_{s \in [0,1]} \left\{ \prod_{j=1}^{r} |g_j(z)| : z = s\sigma_i(\gamma_j) \right\}. \end{aligned}$$

On the straight-line path from 0 to $\sigma_i(\gamma_j)$ we have the bound $\left|z^{p-1}\right| \leq |\sigma_i(\gamma_j)|^{p-1} \leq m_3^{p-1}$. This yields the inequality

$$\begin{aligned} \left| \int_0^{\sigma_i(\gamma_j)} e^{-z} f(z) dz \right| &\leq m_3 m_4 \frac{(b_1 \cdots b_r)^{prn}}{(p-1)!} m_3^{p-1} m_5^p \\ &= m_4 \frac{(b_1 \cdots b_r)^{prn}}{(p-1)!} m_3^p m_5^p. \end{aligned}$$

Combining this with the previous inequality gives

$$\begin{aligned}1 &\leq \left| \sum_{j=1}^{r} \sum_{i=1}^{n} c_j e^{\sigma_i(\gamma_j)} \int_0^{\sigma_i(\gamma_j)} e^{-z} f(z)dz \right| \\ &\leq rnm_1m_2 \left(m_4 \frac{(b_1 \cdots b_r)^{prn}}{(p-1)!} m_3^p m_5^p \right) \\ &= rnm_1m_2m_4 \frac{((b_1 \cdots b_r)^{rn} m_3 m_5)^p}{(p-1)!}.\end{aligned}$$

Since $u^p/(p-1)! \to 0$ as $p \to \infty$, the last term in the inequality above can be made arbitrarily small by choosing p large enough. This gives a contradiction, so our original hypothesis that the exponentials $e^{\alpha_1}, \dots, e^{\alpha_m}$ are linearly dependent over $\mathbb{Q}$ is false. This proves the theorem. □

While we have proved that π and e are transcendental over $\mathbb{Q}$, it is unknown if π is transcendental over $\mathbb{Q}(e)$ or if e is transcendental over $\mathbb{Q}(\pi)$. To discuss this further, we need a definition from Section 19. If K is a field extension of F, then $a_1, \dots, a_n \in K$ are algebraically independent over F if whenever $f \in F[x_1, \dots, x_n]$ is a polynomial with $f(a_1, \dots, a_n) = 0$, then $f = 0$. It is not hard to show that π and e are *algebraically independent* over $\mathbb{Q}$ if and only if π is transcendental over $\mathbb{Q}(e)$, if and only if e is transcendental over $\mathbb{Q}(\pi)$; see Problem 2. A possible generalization of the Lindemann–Weierstrauss theorem is *Schanuel's conjecture*, which states that if $y_1, \dots, y_n$ are $\mathbb{Q}$-linearly independent complex numbers, then at least n of the numbers $y_1, \dots, y_n, e^{y_1}, \dots, e^{y_n}$ are algebraically independent over $\mathbb{Q}$. If Schanuel's conjecture is true, then e and π are algebraically independent over $\mathbb{Q}$; this is left to Problem 3.

Problems

1. The original Lindemann–Weierstrauss theorem states that if $\alpha_1, \dots, \alpha_m$ are $\mathbb{Q}$-linearly independent algebraic numbers, then the exponentials $e^{\alpha_1}, \dots, e^{\alpha_m}$ are algebraically independent; that is, there is no nonzero polynomial $f(x_1, \dots, x_m) \in \mathbb{Q}[x_1, \dots, x_m]$ with $f(e^{\alpha_1}, \dots, e^{\alpha_m}) = 0$. Show that this version of the Lindemann–Weierstrauss theorem is equivalent to the version given in Theorem 14.1.

2. Recall the definition of algebraic independence given at the end of this section. Show that two complex numbers a, b are algebraically independent over $\mathbb{Q}$ if and only if b is transcendental over $\mathbb{Q}(a)$. Conclude that b is transcendental over $\mathbb{Q}(a)$ if and only if a is transcendental over $\mathbb{Q}(b)$.

3. Prove that Schanuel's conjecture implies that there is no nonzero polynomial $f(x,y) \in \mathbb{Q}[x,y]$ with $f(e,\pi) = 0$. In other words, Schanuel's conjecture implies that π and e are algebraically independent over $\mathbb{Q}$.

4. Let
$$\sum_{i=1}^{r} a_i x^{\alpha_i} \quad \text{and} \quad \sum_{i=1}^{s} b_i x^{\beta_i}$$
be functions with a_i, b_i nonzero rational numbers and α_i, β_i algebraic numbers. Assume that the α_i are distinct and that the β_i are distinct. Writing $\sum_j a_j x^{\alpha_j} \cdot \sum_j b_j x^{\beta_j}$ in the form $\sum_k c_k x^{\gamma_j}$ with the γ_j distinct, show that at least one of the c_k is nonzero.

5. Let $a_0, a_1, \ldots, a_n \in \mathbb{Z}$ with $a_n \neq 0$, and let $\{\beta_j\}_{j=1}^{n}$ be the roots of the polynomial $a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$. Let $g(x_1, x_2, \ldots, x_n)$ be a symmetric polynomial in the x_i with integer coefficients. If $t = \deg(g)$, show that $a_n^t g(\beta_1, \ldots, \beta_n)$ is an integer.
(Hint: Use the theorem on symmetric polynomials: If $f(x_1, \ldots, x_n)$ is a symmetric polynomial in the variables $x_1, \ldots, x_n$, then f is a polynomial in the elementary symmetric functions.)

6. Use the infinite series representation $e = \sum_{n=0}^{\infty} 1/n!$ for e to show that e is irrational.
(This approach to proving that e is irrational was found by Fourier.)

7. If u is a nonzero algebraic number, show that $\sin u$ and $\cos u$ are transcendental over $\mathbb{Q}$.

8. If u is a nonzero algebraic number, show that $\tan u$, $\cot u$, $\sec u$, and $\csc u$ are all transcendental over $\mathbb{Q}$.

9. If $u \neq 1$ is a nonzero algebraic number, show that any complex value of $\log u$ is transcendental over $\mathbb{Q}$.

10. If $u \neq 1$ is a nonzero algebraic number and f is any one of the inverse trigonometric functions, show that any complex value of $f(u)$ is transcendental over $\mathbb{Q}$.

11. Let K be the set of all real-valued functions defined and continuous on a dense open subset of $\mathbb{R}$. Define pointwise addition and multiplication of functions $f, g \in K$ in the common domain of f and g.

    (a) Show that K is a field and that K contains the rational function field $\mathbb{R}(x)$.

    (b) Show that the six basic trigonometric functions, $\ln|x|$, and e^x are in K and are not algebraic over $\mathbb{R}(x)$.

12. Early on in the proof of the Lindemann–Weierstrauss theorem, we had an equation $\sum_{j=1}^{m} a_j e^{\alpha_j} = 0$, and we needed equations $\sum_{j=1}^{m} a_j e^{\sigma(\alpha_j)} = 0$, where σ is an automorphism of an appropriate field. If σ is continuous, then we can use infinite series to show that $\sigma(e^a) = e^{\sigma(a)}$. Show that if σ is an automorphism of a subfield F of $\mathbb{C}$, then σ is not continuous unless $\sigma = \text{id}$ or σ is complex conjugation restricted to F.

15 Ruler and Compass Constructions

In the days of the ancient Greeks, some of the major mathematical questions involved constructions with ruler and compass. In spite of the ability of many gifted mathematicians, a number of questions were left unsolved. It was not until the advent of field theory that these questions could be answered. We consider in this section the idea of constructibility by ruler and compass, and we answer the following four classical questions:

1. Is it possible to trisect any angle?
2. Is it possible to double the cube? That is, given a cube of volume V, a side of which can be constructed, is it possible to construct a line segment whose length is that of the side of a cube of volume $2V$?
3. Is it possible to square the circle? That is, given a constructible circle of area A, is it possible to construct a square of area A?
4. For which n is it possible to construct a regular n-gon?

The notion of ruler and compass construction was a theoretical one to the Greeks. A ruler was taken to be an object that could draw perfect, infinitely long lines with no thickness but with no markings to measure distance. The only way to use a ruler was to draw the line passing through two points. Similarly, a compass was taken to be a device that could draw a perfect circle, and the only way it could be used was to draw the circle centered at one point and passing through another. The compass was sometimes referred to as a "collapsible compass"; that is, after drawing a circle, the compass could not be lifted to draw a circle centered at another point with the same radius as that of the previous circle. Likewise, given two points a distance d apart, the ruler cannot be used to mark a point on another line a distance d from a given point on the line.

The assumptions of constructibility are as follows. Two points are given and are taken to be the initial constructible points. Given any two constructible points, the line through these points can be constructed, as can the circle centered at one point passing through the other. A point is constructible if it is the intersection of constructible lines and circles.

The first thing we note is that the collapsibility of the compass is not a problem, nor is not being able to use the ruler to mark distances. Given two constructible points a distance d apart, and a line ℓ with a point P on ℓ, we can construct a point Q on ℓ a distance d from P. Also, if we can construct a circle of radius r, given any constructible point P, we can construct the circle of radius r centered at P. These facts are indicated in Figure 15.1. It is left as an exercise (Problem 4) to describe the construction indicated by the figure.

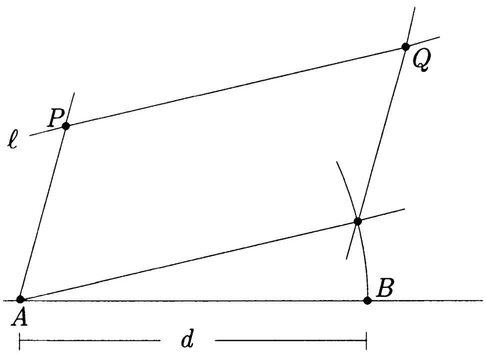


FIGURE 15.1. Construction of Q on ℓ a distance d from P.

There are some standard constructions from elementary geometry that we recall now. Given a line and a point on the line, it is possible to construct a second line through the point perpendicular to the original line. Also, given a line and a point not on the line, it is possible to construct a second line parallel to the original line and passing through the point. These facts are indicated in Figure 15.2.

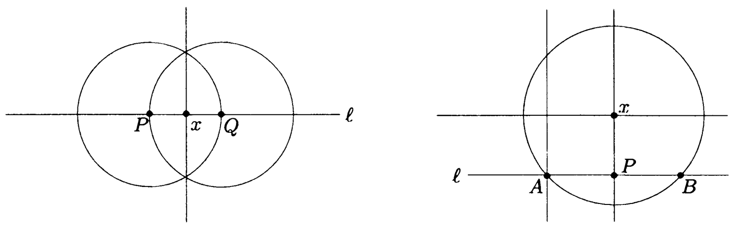


FIGURE 15.2. Construction of lines perpendicular and parallel to ℓ passing through x.

So far, our discussion has been purely geometric. We need to describe ruler and compass constructions algebraically in order to answer our four questions. To do this, we turn to the methods of analytic geometry. Given our original two points, we set up a coordinate system by defining the x-axis to be the line through the points, setting one point to be the origin

and the other to be the point $(1,0)$. We can draw the line perpendicular to the x-axis through the origin to obtain the y-axis.

Let $a \in \mathbb{R}$. We say that a is a constructible number if we can construct two points a distance $|a|$ apart. Equivalently, a is constructible if we can construct either of the points $(a,0)$ or $(0,a)$. If a and b are constructible numbers, elementary geometry tells us that $a+b$, $a-b$, ab, and a/b (if $b \neq 0$) are all constructible. Therefore, the set of all constructible numbers is a subfield of $\mathbb{R}$. Furthermore, if $a > 0$ is constructible, then so is $\sqrt{a}$. These facts are illustrated in Figures 15.3–15.5.



FIGURE 15.3. Construction of $a+b$ and $a-b$.

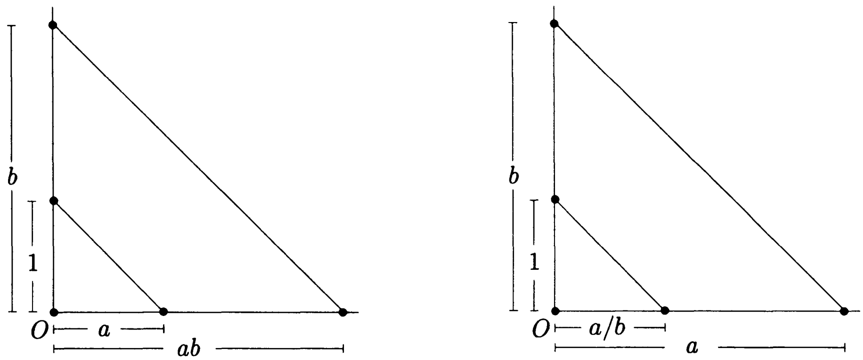


FIGURE 15.4. Construction of ab and a/b.

Suppose that P is a constructible point, and set $P = (a,b)$ in our coordinate system. We can construct the lines through P perpendicular to the x-axis and y-axis; hence, we can construct the points $(a,0)$ and $(0,b)$. Therefore, a and b are constructible numbers. Conversely, if a and b are constructible numbers, we can construct $(a,0)$ and $(0,b)$, so we can construct P as the intersection of the line through $(a,0)$ parallel to the y-axis with the line through $(0,b)$ parallel to the x-axis. Thus, $P=(a,b)$ is constructible if and only if a and b are constructible numbers.

In order to construct a number c, we must draw a finite number of lines and circles in such a way that $|c|$ is the distance between two points of intersection. Equivalently, we must draw lines and circles so that $(c,0)$ is a point of intersection. If we let K be the field generated over $\mathbb{Q}$ by all the numbers obtained in some such construction, we obtain a subfield of the field of constructible numbers. To give a criterion for when a number

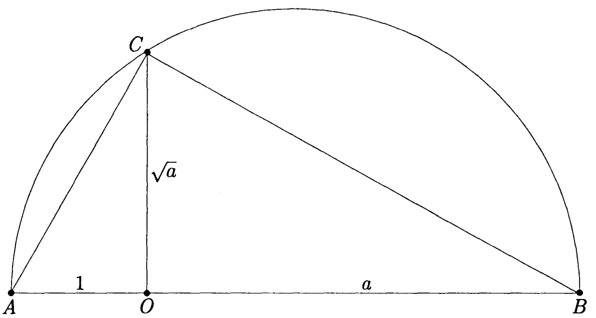


FIGURE 15.5. Construction of $\sqrt{a}$.

is constructible, we need to relate constructibility to properties of the field extension $K/\mathbb{Q}$. We do this with analytic geometry. Let K be a subfield of $\mathbb{R}$. Given any two points in the plane of K, we obtain a line through these points. This will be called a *line in* K. It is not hard to show that a line in K has an equation of the form $ax + by + c = 0$ with $a, b, c \in K$. If P and Q are points in the plane of K, the circle with center P passing through Q is called a *circle in* K. Again, it is not hard to show that the equation of a circle in K can be written in the form $x^2 + y^2 + ax + by + c = 0$ for some $a, b, c \in K$. The next lemma gives us a connection between constructibility and field extensions.

Lemma 15.1 *Let K be a subfield of $\mathbb{R}$.*

1. *The intersection of two lines in K is either empty or is a point in the plane of K.*

2. *The intersection of a line and a circle in K is either empty or consists of one or two points in the plane of $K(\sqrt{u})$ for some $u \in K$ with $u \geq 0$.*

3. *The intersection of two circles in K is either empty or consists of one or two points in the plane of $K(\sqrt{u})$ for some $u \in K$ with $u \geq 0$.*

Proof. The first statement is an easy calculation. For the remaining two statements, it suffices to prove statement 2, since if $x^2+y^2+ax+by+c = 0$ and $x^2 + y^2 + a'x + b'y + c' = 0$ are the equations of circles C and C', respectively, then their intersection is the intersection of C with the line $(a - a')x + (b - b')y + (c - c') = 0$. So, to prove statement 2, suppose that our line L in K has the equation $dx + ey + f = 0$. We assume that $d \neq 0$, since if $d = 0$, then $e \neq 0$. By dividing by d, we may then assume that $d = 1$. Plugging $-x = ey + f$ into the equation of C, we obtain

$$(e^2 + 1)y^2 + (2ef - ae + b)y + (f^2 - af + c) = 0.$$

Writing this equation in the form $\alpha y^2 + \beta y + \gamma = 0$, if $\alpha = 0$, then $y \in K$. If $\alpha \neq 0$, then completing the square shows that either $L \cap C = \varnothing$ or $y \in K(\sqrt{\beta^2 - 4\alpha\gamma})$ with $\beta^2 - 4\alpha\gamma \geq 0$. □

From this lemma, we can turn the definition of constructibility into a property of field extensions of $\mathbb{Q}$, and in doing so obtain a criterion for when a number is constructible.

Theorem 15.2 *A real number c is constructible if and only if there is a tower of fields $\mathbb{Q} = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_r$ such that $c \in K_r$ and $[K_{i+1} : K_i] \leq 2$ for each i. Therefore, if c is constructible, then c is algebraic over $\mathbb{Q}$, and $[\mathbb{Q}(c) : \mathbb{Q}]$ is a power of 2.*

Proof. If c is constructible, then the point $(c, 0)$ can be obtained from a finite sequence of constructions starting from the plane of $\mathbb{Q}$. We then obtain a finite sequence of points, each an intersection of constructible lines and circles, ending at $(c, 0)$. By Lemma 15.1, the first point either lies in $\mathbb{Q}$ or in $\mathbb{Q}(\sqrt{u})$ for some u. This extension has degree either 1 or 2. Each time we construct a new point, we obtain a field extension whose degree over the previous field is either 1 or 2 by the lemma. Thus, we obtain a sequence of fields

$$\mathbb{Q} = K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_r$$

with $[K_{i+1} : K_i] \leq 2$ and $c \in K_r$. Therefore, $[K_r : \mathbb{Q}] = 2^n$ for some n. However, $[\mathbb{Q}(c) : \mathbb{Q}]$ divides $[K_r : \mathbb{Q}]$, so $[\mathbb{Q}(c) : \mathbb{Q}]$ is also a power of 2.

For the converse, suppose that we have a tower $\mathbb{Q} = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_r$ with $c \in K_r$ and $[K_{i+1} : K_i] \leq 2$ for each i. We show that c is constructible by induction on r. If $r = 0$, then $c \in \mathbb{Q}$, so c is constructible. Assume then that $r > 0$ and that elements of K_{r-1} are constructible. Since $[K_r : K_{r-1}] \leq 2$, the quadratic formula shows that we may write $K_r = K_{r-1}(\sqrt{a})$ for some $a \in K_{r-1}$. Since a is constructible by assumption, so is $\sqrt{a}$. Therefore, $K_r = K_{r-1}(\sqrt{a})$ lies in the field of constructible numbers; hence, c is constructible. □

With this theorem, we are now able to answer the four questions posed earlier. We first consider trisection of angles. An angle of measure θ is constructible if we can construct two intersecting lines such that the angle between them is θ. For example, a 60° angle can be constructed because the point $(\sqrt{3}/2, 1/2)$ is constructible, and the line through this point and $(0, 0)$ makes an angle of 60° with the x-axis. Suppose that P is the point of intersection on two constructible lines. By drawing a circle of radius 1 centered at P, Figure 15.6 shows that if θ is the angle between the two lines, then $\sin\theta$ and $\cos\theta$ are constructible numbers. Conversely, if $\sin\theta$ and $\cos\theta$ are constructible, then θ is a constructible angle (see Problem 2). In order to trisect an angle of measure θ, we would need to be able to construct an angle of $\theta/3$.

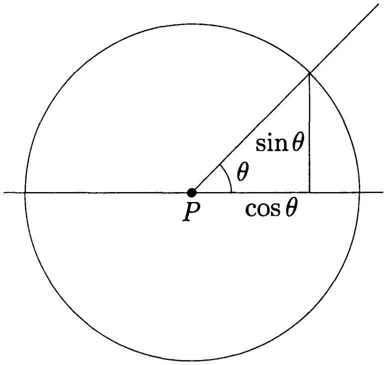


FIGURE 15.6. Construction of sines and cosines.

Theorem 15.3 *It is impossible to trisect a* 60° *angle by ruler and compass construction.*

Proof. As noted above, a 60° angle can be constructed. If a 60° angle can be trisected, then it is possible to construct the number $\alpha = \cos 20^\circ$. However, the triple angle formula $\cos 3\theta = 4\cos^3\theta - 3\cos\theta$ gives $4\alpha^3 - 3\alpha = \cos 60^\circ = 1/2$. Thus, α is algebraic over $\mathbb{Q}$. The polynomial $8x^3 - 6x - 1$ is irreducible over $\mathbb{Q}$ because it has no rational roots. Therefore, $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 3$, so α is not constructible. A 20° angle cannot then be constructed, so a 60° degree angle cannot be trisected. □

This theorem does not say that no angle can be trisected. A 90° angle can be trisected, since a 30° angle can be constructed. This theorem only says that not all angles can be trisected, so there is no method that will trisect an arbitrary angle.

The second classical impossibility we consider is the doubling of a cube.

Theorem 15.4 *It is impossible to double a cube of length* 1 *by ruler and compass construction.*

Proof. The length of a side of a cube of volume 2 is $\sqrt[3]{2}$. The minimal polynomial of $\sqrt[3]{2}$ over $\mathbb{Q}$ is $x^3 - 2$. Thus, $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$ is not a power of 2, so $\sqrt[3]{2}$ is not constructible. □

The third of the classical impossibilities is the squaring of a circle. For this, we need to use the fact that π is transcendental over $\mathbb{Q}$.

Theorem 15.5 *It is impossible to square a circle of radius* 1.

Proof. We are asking whether we can construct a square of area π. To do so requires us to construct a line segment of length $\sqrt{\pi}$, which is impossible since $\sqrt{\pi}$ is transcendental over $\mathbb{Q}$ by the Lindemann–Weierstrauss theorem; hence, $\sqrt{\pi}$ is not algebraic of degree a power of 2. □

Our last question concerns construction of regular n-gons. To determine which regular n-gons can be constructed, we will need information about cyclotomic extensions. Recall from Section 7 that if ω is a primitive nth root of unity, then $[\mathbb{Q}(\omega) : \mathbb{Q}] = \phi(n)$, where ϕ is the Euler phi function.

Theorem 15.6 *A regular n-gon is constructible if and only if $\phi(n)$ is a power of* 2.

Proof. We point out that a regular n-gon is constructible if and only if the central angles $2\pi/n$ are constructible, and this occurs if and only if $\cos(2\pi/n)$ is a constructible number. Let $\omega = e^{2\pi i/n} = \cos(2\pi/n) + i\sin(2\pi/n)$, a primitive nth root of unity. Then $\cos(2\pi/n) = \frac{1}{2}(\omega + \omega^{-1})$, since $\omega^{-1} = \cos(2\pi/n) - i\sin(2\pi/n)$. Thus, $\cos(2\pi/n) \in \mathbb{Q}(\omega)$. However, $\cos(2\pi/n) \in \mathbb{R}$ and $\omega \notin \mathbb{R}$, so $\mathbb{Q}(\omega) \neq \mathbb{Q}(\cos(2\pi/n))$. But ω is a root of $x^2 - 2\cos(2\pi/n)x + 1$, as an easy calculation shows, so $[\mathbb{Q}(\omega) : \mathbb{Q}(\cos(2\pi/n))] = 2$. Therefore, if $\cos(2\pi/n)$ is constructible, then $[\mathbb{Q}(\cos(2\pi/n)) : \mathbb{Q}]$ is a power of 2. Hence, $\phi(n) = [\mathbb{Q}(\omega) : \mathbb{Q}]$ is also a power of 2.

Conversely, suppose that $\phi(n)$ is a power of 2. The field $\mathbb{Q}(\omega)$ is a Galois extension of $\mathbb{Q}$ with Abelian Galois group by Proposition 7.2. If $H = \text{Gal}(\mathbb{Q}(\omega)/\mathbb{Q}(\cos(2\pi/n)))$, by the theory of finite Abelian groups there is a chain of subgroups

$$H_0 \subseteq H_1 \subseteq \cdots \subseteq H_r = H$$

with $|H_{i+1} : H_i| = 2$. If $L_i = \mathcal{F}(H_i)$, then $[L_i : L_{i+1}] = 2$; thus, $L_i = L_{i+1}(\sqrt{u_i})$ for some u_i. Since $L_i \subseteq \mathbb{Q}(\cos(2\pi/n)) \subseteq \mathbb{R}$, each of the $u_i \geq 0$. Since the square root of a constructible number is constructible, we see that everything in $\mathbb{Q}(\cos(2\pi/n))$ is constructible. Thus, $\cos(2\pi/n)$ is constructible, so a regular n-gon is constructible. □

This theorem shows, for example, that a regular 9-gon is not constructible and a regular 17-gon is constructible. An explicit algorithm for constructing a regular 17-gon was given by Gauss in 1801. If $n = p_1^{m_1} \cdots p_r^{m_r}$ is the prime factorization of n, then $\phi(n) = \prod_i p^{m_i - 1}(p_i - 1)$. Therefore, $\phi(n)$ is a power of 2 if and only if $n = 2^s q_1 \cdots q_r$, where $r, s \geq 0$, and the q_i are primes of the form $2^m + 1$. In order to determine which regular n-gons are constructible, it then reduces to determining the primes of the form $2^m + 1$.

Problems

1. Use the figures in this section to describe how to construct $a+b$, $a-b$, ab, a/b, and $\sqrt{a}$, provided that a and b are constructible.

2. If $\sin\theta$ and $\cos\theta$ are constructible numbers, show that θ is a constructible angle.

3. If an angle θ can be constructed, show that a line passing through the origin can be constructed such that the angle between this line and the x-axis is θ.

4. Use the figures of this section to answer the following questions.

   (a) Given two points a distance d apart and a constructible point P on a line ℓ, show that it is possible to construct a point Q on ℓ a distance d from P.

   (b) Given that some circle of radius r can be constructed, if P is a constructible point, show that the circle of radius r centered at P can be constructed.

   (c) Given a line ℓ and a point P on ℓ, show that it is possible to construct the line through P perpendicular to ℓ.

   (d) Given a line ℓ and a point P not on ℓ, show that it is possible to construct the line through P parallel to ℓ.

5. Let $c \in \mathbb{R}$ be a root of an irreducible quartic over $\mathbb{Q}$. Let N be the normal closure of $\mathbb{Q}(c)/\mathbb{Q}$.

   (a) If $\mathrm{Gal}(N/\mathbb{Q})$ is isomorphic to either D_4 or a group of order 4, show that c is constructible.

   (b) If $\mathrm{Gal}(N/\mathbb{Q})$ is isomorphic to either A_4 or S_4, show that c is not constructible.

6. Let $c \in \mathbb{R}$ be algebraic over $\mathbb{Q}$, and let N be the normal closure of $\mathbb{Q}(c)/\mathbb{Q}$. If $[N : \mathbb{Q}]$ is a power of 2, show that c is constructible.

7. This problem gives a partial converse to Theorem 15.2. If $c \in \mathbb{R}$ is algebraic over $\mathbb{Q}$ and if N is the normal closure of $\mathbb{Q}(c)/\mathbb{Q}$, then show that c is constructible if and only if $[N : \mathbb{Q}]$ is a power of 2.
   (The criterion for constructibility proven in this section is much like the definition of solvable by radicals given in Section 16. If you work this problem, some proofs of the next section will be easier to understand.)

8. A Fermat number is a number of the form $2^{2^r} + 1$ for some r. Suppose that p is an odd prime such that a regular p-gon is constructible. Show that p is a Fermat number.

16 Solvability by Radicals

In this section, we address one of the driving forces of mathematics for hundreds of years, the solvability of polynomial equations. As we saw in Section

13, formulas for the roots of cubic and quartic polynomials are known and had been found by the mid-sixteenth century. While it was over a thousand years between the discovery of the quadratic formula and the solution of the cubic, the solution of the quartic came soon after the solution of the cubic. This success led mathematicians to believe that formulas for the roots of polynomials of arbitrary degree could be found. However, nothing had been discovered for polynomials of higher degree until Abel proved in a paper published in 1824 that there is no "algebraic" solution of the quintic; that is, there is no solution that expresses the roots in terms of the coefficients, arithmetic operations, and radicals. The full story of solvability of polynomials was then discovered by Galois, who proved a necessary and sufficient condition for a polynomial to be solvable. His work introduced the notion of a group and was the birth of abstract algebra.

We need to make precise what it means for a polynomial to be solvable. Consider, for example, the polynomial x^4-6x^2+7. Its roots are $\pm\sqrt{3 \pm \sqrt{2}}$, all of which lie in the extension $\mathbb{Q}(\sqrt{2}, \sqrt{3+\sqrt{2}}, \sqrt{3-\sqrt{2}})$ of $\mathbb{Q}$. This extension gives rise to the chain of simple extensions

$$\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(\sqrt{2})(\sqrt{3+\sqrt{2}}) \subseteq \mathbb{Q}(\sqrt{2}, \sqrt{3+\sqrt{2}})(\sqrt{3-\sqrt{2}}),$$

where each successive field is obtained from the previous one by adjoining the root of an element of the previous field. This example motivates the following definitions.

Definition 16.1 *A field extension K of F is a radical extension if $K = F(a_1, \ldots, a_r)$, such that there are integers $n_1, \ldots, n_r$ with $a_1^{n_1} \in F$ and $a_i^{n_i} \in F(a_1, \ldots, a_{i-1})$ for all $i > 1$. If $n_1 = \cdots = n_r = n$, then K is called an n-radical extension of F.*

Definition 16.2 *If $f(x) \in F[x]$, then f is solvable by radicals if there is a radical extension L/F such that f splits over L.*

If K and F are as in the first definition, then K is an n-radical extension of F for $n = n_1 \cdots n_r$ since $a_i^n \in F(a_1, \ldots, a_{i-1})$ for each i. The definition of radical extension is equivalent to the following statement: K is a radical extension of F if there is a chain of fields

$$F = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_r = K,$$

where $F_{i+1} = F_i(a_i)$ for some $a_i \in F_{i+1}$ with $a_i^{n_i} \in F_i$ for each i. From the definition, it follows easily that if K/F is a radical extension and L/K is a radical extension, then L/F is a radical extension.

Example 16.3 Any 2-Kummer extension of a field F of characteristic not 2 is a 2-radical extension of F by Theorem 11.4. Also, if K/F is a cyclic extension of degree n, and if F contains a primitive nth root of unity, then K is an n-radical extension of F by Theorem 9.5.

Example 16.4 If $K = \mathbb{Q}(\sqrt[4]{2})$, then K is both a 4-radical extension and a 2-radical extension of $\mathbb{Q}$. The second statement is true by considering the tower

$$\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(\sqrt{2})(\sqrt{\sqrt{2}}) = \mathbb{Q}(\sqrt[4]{2}).$$

Example 16.5 Let $c \in \mathbb{R}$. By Theorem 15.2, c is constructible if and only if there is a tower $\mathbb{Q} = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_r$ such that for each i, $F_{i+1} = F_i(\sqrt{a_i})$ for some $a_i \in F_i$, and $c \in F_r$. Therefore, c is constructible if and only if c lies in a subfield K of $\mathbb{R}$ such that K is a 2-radical extension of $\mathbb{Q}$.

The definition of solvability by radicals does not say that the splitting field of f over F is itself a radical extension. It is possible for f to be solvable by radicals but that its splitting field over F is not a radical extension. However, if F contains "enough" roots of unity, then the splitting field of a solvable polynomial is a radical extension of F. For an example of the first statement, see Example 16.13. The second statement is addressed in Problem 3.

The next lemma is the key technical piece of the proof of the characterization of solvability by radicals.

Lemma 16.6 *Let K be an n-radical extension of F, and let N be the normal closure of K/F. Then N is an n-radical extension of F.*

Proof. Let $K = F(\alpha_1, \ldots, \alpha_r)$ with $\alpha_i^n \in F(\alpha_1, \ldots, \alpha_{i-1})$. We argue by induction on r. If $r = 1$, then $K = F(\alpha)$ with $\alpha^n = a \in F$. Then $N = F(\beta_1, \ldots, \beta_m)$, where the β_i are the roots of $\min(F, \alpha)$. However, this minimal polynomial divides $x^n - a$, so $\beta_i^n = a$. Thus, N is an n-radical extension of F. Now suppose that $r > 1$. Let N_0 be the normal closure of $F(\alpha_1, \ldots, \alpha_{r-1})$ over F. By induction, N_0 is an n-radical extension of F. Since N_0 is the splitting field over F of $\{\min(F, \alpha_i) : 1 \le i \le r-1\}$, and N is the splitting field of all $\min(F, \alpha_i)$, we have $N = N_0(\gamma_1, \ldots, \gamma_m)$, where the γ_i are roots of $\min(F, \alpha_r)$. Also, $\alpha_r^n = b$ for some $b \in F(\alpha_1, \ldots, \alpha_{r-1}) \subseteq N_0$. By the isomorphism extension theorem, for each i there is a $\sigma_i \in \mathrm{Gal}(N/F)$ with $\sigma_i(\alpha_r) = \gamma_i$. Therefore, $\gamma_i^n = \sigma_i(b)$ by Proposition 3.28. However, N_0 is normal over F, and $b \in N_0$, so $\sigma_i(b) \in N_0$. Thus, each γ_i is an nth power of some element of N_0, so N is an n-radical extension of N_0. Since N_0 is an n-radical extension of F, we see that N is an n-radical extension of F. □

We need some group theory in order to state and prove Galois' theorem on solvability by radicals. The key group theoretic notion is that of solvability of a group. A little more information on solvability can be found in Appendix C.

Definition 16.7 *A group G is solvable if there is a chain of subgroups*

$$\langle e\rangle = H_0 \subseteq H_1 \subseteq \cdots \subseteq H_n = G$$

such that for all i, the subgroup H_i is normal in H_{i+1} and the quotient group H_{i+1}/H_i is Abelian.

The following two propositions are the facts that we require about solvability. The first is proved in Appendix C, and the second can be found in any good group theory book.

Proposition 16.8 *Let G be a group and N be a normal subgroup of G. Then G is solvable if and only if N and G/N are solvable.*

Proposition 16.9 *If $n \geq 5$, then S_n is not solvable.*

We now prove Galois' theorem characterizing polynomials that are solvable by radicals.

Theorem 16.10 (Galois) *Let $\text{char}(F) = 0$ and let $f(x) \in F[x]$. If K is a splitting field of f over F, then f is solvable by radicals if and only if $\text{Gal}(K/F)$ is a solvable group.*

Proof. Suppose that f is solvable by radicals. Then there is an n-radical extension M/F with $K \subseteq M$. Let ω be a primitive nth root of unity in some extension field of M. The existence of ω follows from the assumption that $\text{char}(F) = 0$. Then $M(\omega)/M$ is an n-radical extension, so $M(\omega)/F$ is an n-radical extension. Let L be the normal closure of $M(\omega)/F$. By Lemma 16.6, L is an n-radical extension of F. Thus, L is also an n-radical extension of $F(\omega)$. Therefore, there is a sequence of fields

$$F = F_0 \subseteq F_1 = F(\omega) \subseteq F_2 \subseteq \cdots \subseteq F_r = L,$$

where $F_{i+1} = F_i(\alpha_i)$ with $\alpha_i^n \in F_i$. For $i \geq 1$, the extension F_{i+1}/F_i is Galois with a cyclic Galois group by Theorem 9.6, since F_i contains a primitive nth root of unity. Also, F_1/F_0 is an Abelian Galois extension, since F_1 is a cyclotomic extension of F. Because $\text{char}(F) = 0$ and L/F is normal, L/F is Galois by Theorem 4.9. Let $G = \text{Gal}(L/F)$ and $H_i = \text{Gal}(L/F_i)$. We have the chain of subgroups

$$\langle \text{id}\rangle = H_r \subseteq H_{r-1} \subseteq \cdots \subseteq H_0 = G.$$

By the fundamental theorem, H_{i+1} is normal in H_i since F_{i+1} is Galois over F_i. Furthermore, $H_i/H_{i+1} \cong \text{Gal}(F_{i+1}/F_i)$, so H_i/H_{i+1} is an Abelian group. Thus, we see that G is solvable, so $\text{Gal}(K/F)$ is also solvable, since $\text{Gal}(K/F) \cong G/\text{Gal}(L/K)$.

For the converse, suppose that $\mathrm{Gal}(K/F)$ is a solvable group. We have a chain

$$\mathrm{Gal}(K/F) = H_0 \supseteq H_1 \supseteq \cdots \supseteq H_r = \langle \mathrm{id} \rangle$$

with H_{i+1} normal in H_i and H_i/H_{i+1} Abelian. Let $K_i = \mathcal{F}(H_i)$. By the fundamental theorem, K_{i+1} is Galois over K_i and $\mathrm{Gal}(K_{i+1}/K_i) \cong H_i/H_{i+1}$. Let n be the exponent of $\mathrm{Gal}(K/F)$, let ω be a primitive nth root of unity, and set $L_i = K_i(\omega)$. We have the chain of fields

$$F \subseteq L_0 \subseteq L_1 \subseteq \cdots \subseteq L_r$$

with $K \subseteq L_r$. Note that $L_{i+1} = L_i K_{i+1}$. Since K_{i+1}/K_i is Galois, by the theorem of natural irrationalities, L_{i+1}/L_i is Galois and $\mathrm{Gal}(L_{i+1}/L_i)$ is isomorphic to a subgroup of $\mathrm{Gal}(K_{i+1}/K_i)$. This second group is isomorphic to H_i/H_{i+1}, an Abelian group. Thus, $\mathrm{Gal}(L_{i+1}/L_i)$ is Abelian, and its exponent divides n. The field L_{i+1} is an n-Kummer extension of L_i by Theorem 11.4, so L_{i+1} is an n-radical extension of L_i. Since $L_0 = F(\omega)$ is a radical extension, transitivity shows that L_r is a radical extension of F. As $K \subseteq L_r$, the polynomial f is solvable by radicals. □

Our definition of radical extension is somewhat lacking for fields of characteristic p, in that Theorem 16.10 is not true in general for prime characteristic. However, by modifying the definition of radical extension in an appropriate way, we can extend this theorem to fields of characteristic p. This is addressed in Problem 2. Also, note that we only needed that $\mathrm{char}(F)$ does not divide n in both directions of the proof. Therefore, the proof above works for fields of characteristic p for adequately large p.

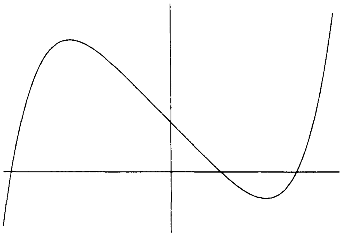
Let k be a field. The general nth degree polynomial over k is the polynomial

$$\begin{aligned} f(x) &= (x - t_1)(x - t_2)\cdots(x - t_n) = x^n - s_1 x^{n-1} + \cdots + (-1)^n s_n \\ &\in k(t_1, \ldots, t_n)[x], \end{aligned}$$

where the s_i are the elementary symmetric functions in the t_i. If we could find a formula for the roots of f in terms of the coefficients of f, we could use this to find a formula for the roots of an arbitrary nth degree polynomial over k. If $n \le 4$, we found formulas for the roots of f in Section 13. For $n \ge 5$, the story is different. The symmetric group S_n is a group of automorphisms on $K = k(t_1, \ldots, t_n)$ as in Example 2.22, and the fixed field is $F = k(s_1, \ldots, s_n)$. Therefore, $\mathrm{Gal}(K/F) = S_n$. Theorem 16.10 shows that no such formula exists if $n \ge 5$.

Corollary 16.11 *Let $f(x)$ be the general nth degree polynomial over a field of characteristic 0. If $n \ge 5$, then f is not solvable by radicals.*

Example 16.12 Let $f(x) = x^5 - 4x + 2 \in \mathbb{Q}[x]$. By graphing techniques of calculus, we see that this polynomial has exactly two nonreal roots, as indicated in the graph below.



Furthermore, f is irreducible over $\mathbb{Q}$ by the Eisenstein criterion. Let K be the splitting field of f over $\mathbb{Q}$. Then $[K : \mathbb{Q}]$ is a multiple of 5, since any root of f generates a field of dimension 5 over $\mathbb{Q}$. Let $G = \text{Gal}(K/\mathbb{Q})$. We can view $G \subseteq S_5$. There is an element of G of order 5 by Cayley's theorem, since 5 divides $|G|$. Any element of S_5 of order 5 is a 5-cycle. Also, if σ is complex conjugation restricted to K, then σ permutes the two nonreal roots of f and fixes the three others, so σ is a transposition. The subgroup of S_5 generated by a transposition and a 5-cycle is all of S_5, so $G = S_5$ is not solvable. Thus, f is not solvable by radicals.

Example 16.13 Let $f(x) = x^3 - 3x + 1 \in \mathbb{Q}[x]$, and let K be the splitting field of f over $\mathbb{Q}$. We show that f is solvable by radicals but that K is not a radical extension of $\mathbb{Q}$. Since f has no roots in $\mathbb{Q}$ and $\deg(f) = 3$, the polynomial f is irreducible over $\mathbb{Q}$. The discriminant of f is $81 = 9^2$, so the Galois group of $K/\mathbb{Q}$ is A_3 and $[K : \mathbb{Q}] = 3$, by Corollary 12.4. Therefore, $\text{Gal}(K/F)$ is solvable, so f is solvable by radicals by Galois' theorem. If K is a radical extension of $\mathbb{Q}$, then there is a chain of fields

$$\mathbb{Q} \subseteq F_1 \subseteq \cdots \subseteq F_r = K$$

with $F_i = F_{i-1}(\alpha_i)$ and $\alpha_i^n \in F_{i-1}$ for some n. Since $[K : \mathbb{Q}]$ is prime, we see that there is only one proper inclusion in this chain. Thus, $K = \mathbb{Q}(b)$ with $b^n = u \in \mathbb{Q}$ for some n. The minimal polynomial $p(x)$ of b over $\mathbb{Q}$ splits in K, since $K/\mathbb{Q}$ is normal. Let b' be another root of $p(x)$. Then $b^n = (b')^n = u$, so b'/b is an nth root of unity. Suppose that $\mu = b'/b$ is a primitive mth root of unity, where m divides n. Then $\mathbb{Q}(\mu) \subseteq K$, so $[\mathbb{Q}(\mu) : \mathbb{Q}] = \phi(m)$ is either 1 or 3. An easy calculation shows that $\phi(m) \neq 3$ for all m. Thus, $[\mathbb{Q}(\mu) : \mathbb{Q}] = 1$, so $\mu \in \mathbb{Q}$. However, the only roots of unity in $\mathbb{Q}$ are ± 1, so $\mu = \pm 1$. Therefore $b' = \pm b$. This proves that $p(x)$ has at most two roots, so $[\mathbb{Q}(b) : \mathbb{Q}] \leq 2 < [K : \mathbb{Q}]$, a contradiction to the equality $\mathbb{Q}(b) = K$. Thus, K is not a radical extension of $\mathbb{Q}$.

Problems

1. Let M be an algebraic closure of $\mathbb{F}_p$, and let $F = M(x)$. Show that $f(t) = t^p - t - x$ is not solvable by radicals over F but that the Galois group of the splitting field of f over F is cyclic.

2. Let F be a field of characteristic $p > 0$. Extend the definition of radical extension as follows. An extension K of F is a radical extension if there is a chain of fields $F = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_n = K$ such that $F_{i+1} = F_i(u_i)$ for some u_i, with either $u_i^{n_i} \in F_i$ for some n_i, or $u_i^p - u_i \in F_i$. Prove that Theorem 16.10 holds in prime characteristic with this definition of radical extension.

3. Let $f(x) \in F[x]$ be solvable by radicals. If F contains a primitive nth root of unity for all n, show that the splitting field of f over F is a radical extension of F. After working through this, figure out just which roots of unity F needs to have for the argument to work.

4. *Solvability by real radicals.* Suppose that $f(x) \in \mathbb{Q}[x]$ has all real roots. If f is solvable by radicals, is f solvable by "real radicals"? That is, does there exist a chain of fields $\mathbb{Q} = Q_0 \subseteq Q_1 \subseteq \cdots \subseteq Q_n \subseteq \mathbb{R}$ such that Q_n contains all the roots of f, and $Q_{i+1} = Q_i(\sqrt[n_i]{a_i})$? The answer is no, in general, and this problem gives a criterion for when f is solvable by real radicals. Use the following steps to prove the following statement: If $f(x) \in \mathbb{Q}[x]$ is an irreducible polynomial with all real roots, and if N is the splitting field of f over $\mathbb{Q}$, then $[N : \mathbb{Q}]$ is a power of 2 if and only if f is solvable by real radicals. You may assume the following nontrivial fact: If $F \subseteq K$ are subfields of $\mathbb{R}$ with $K = F(a)$ such that $a^n \in F$, and if L is an intermediate field of K/F Galois over F, then $[L : F] \le 2$.

   (a) If $[N : \mathbb{Q}]$ is not a power of 2, let p be an odd prime divisor of $[N : \mathbb{Q}]$. Let P be the subgroup of $G = \mathrm{Gal}(N/\mathbb{Q})$ generated by all elements of order p. Show that P is a normal subgroup of G and that $P \neq \langle \mathrm{id} \rangle$.

   (b) Let α be a root of f, and let $T = \mathbb{Q}(\alpha)$. If $H = \mathrm{Gal}(N/T)$, show that P is not contained in H. Conclude that there is an element $\sigma \in G$ of order p not contained in H.

   (c) Let $F = \mathcal{F}(\langle \sigma \rangle)$. Show that $\alpha \notin F$. Let Q_i be in the chain above, and set $F_i = FQ_i$. Show that there is an integer $r > 0$ with $\alpha \notin F_{r-1}$ but $\alpha \in F_r$. Show that $F = F_{r-1} \cap N$ and $N \subseteq F_r$.

   (d) Let $E = NF_{r-1}$. Then $F_{r-1} \subseteq E \subseteq F_r$. Conclude from the assumption above and the theorem of natural irrationalities that $p = [E : F_{r-1}] \le 2$, a contradiction.
   (A full proof of this criterion for solvability by real radicals can be found in Isaacs [14].)

IV
Infinite Algebraic Extensions

In this chapter, we investigate infinite Galois extensions and prove an analog of the fundamental theorem of Galois theory for infinite extensions. The key idea is to put a topology on the Galois group of an infinite dimensional Galois extension and then use this topology to determine which subgroups of the Galois group arise as Galois groups of intermediate extensions. We also give a number of constructions of infinite Galois extensions, constructions that arise in quadratic form theory, number theory, and Galois cohomology, among other places.

17 Infinite Galois Extensions

In this section, we consider Galois extensions K/F of arbitrary degree and prove a fundamental theorem for such extensions. If $[K : F] = \infty$, then not all subgroups of $\mathrm{Gal}(K/F)$ have the form $\mathrm{Gal}(K/L)$ for some intermediate extension L (see Problem 4). We need more information about $\mathrm{Gal}(K/F)$ in order to determine when a subgroup is of the form $\mathrm{Gal}(K/L)$. It turns out that the right way to look at $\mathrm{Gal}(K/F)$ is to put a topology on it. This was first done by Krull in the 1920s, and we see below that the subgroups of $\mathrm{Gal}(K/F)$ of the form $\mathrm{Gal}(K/L)$ are precisely the subgroups that are closed with respect to the topology we define on $\mathrm{Gal}(K/F)$. We assume in this section that the reader is familiar with the basic ideas of point set topology, in particular with the notions of compactness and the Hausdorff

property. The interested reader can find a discussion of these notions in Appendix E.

Let K be a Galois extension of F. We will use the following notation for the rest of this section. Let

$$G = \mathrm{Gal}(K/F),$$
$$\mathcal{I} = \{\, E : F \subseteq E \subseteq K,\ [E : F] < \infty \text{ and } E/F \text{ is Galois} \,\},$$
$$\mathcal{N} = \{\, N \subseteq G : N = \mathrm{Gal}(K/E) \text{ for some } E \in \mathcal{I} \,\}.$$

Recall part 3 of Proposition 3.28: If K/F is normal, and if $F \subseteq L \subseteq K \subseteq N$ are fields with $\tau : L \longrightarrow N$ an F-homomorphism, then $\tau(L) \subseteq K$, and there is a $\sigma \in \mathrm{Gal}(K/F)$ with $\sigma|_L = \tau$. We will use this result frequently.

We start off by proving a few simple properties of the sets $\mathcal{I}$ and $\mathcal{N}$.

Lemma 17.1 *If $\alpha_1, \ldots, \alpha_n \in K$, then there is an $E \in \mathcal{I}$ with $\alpha_i \in E$ for all i.*

Proof. Let $E \subseteq K$ be the splitting field of the minimal polynomials of the α_i over F. Then, as each α_i is separable over F, the field E is normal and separable over F; hence, E is Galois over F. Since there are finitely many α_i, we have $[E : F] < \infty$, so $E \in \mathcal{I}$. □

Lemma 17.2 *Let $N \in \mathcal{N}$, and set $N = \mathrm{Gal}(K/E)$ with $E \in \mathcal{I}$. Then $E = \mathcal{F}(N)$ and N is normal in G. Moreover, $G/N \cong \mathrm{Gal}(E/F)$. Thus, $|G/N| = |\,\mathrm{Gal}(E/F)| = [E : F] < \infty$.*

Proof. Since K is normal and separable over F, the field K is also normal and separable over E, so K is Galois over E. Therefore, $E = \mathcal{F}(N)$. As in the proof of the fundamental theorem, the map $\theta : G \longrightarrow \mathrm{Gal}(E/F)$ given by $\sigma \mapsto \sigma|_E$ is a group homomorphism with kernel $\mathrm{Gal}(K/E) = N$. Proposition 3.28 shows that θ is surjective. The remaining statements then follow. □

Lemma 17.3 *We have $\bigcap_{N \in \mathcal{N}} N = \{\mathrm{id}\}$. Furthermore, $\bigcap_{N \in \mathcal{N}} \sigma N = \{\sigma\}$ for all $\sigma \in G$.*

Proof. Let $\tau \in \bigcap_{N \in \mathcal{N}} N$ and let $a \in K$. By Lemma 17.1, there is an $E \in \mathcal{I}$ with $a \in E$. Set $N = \mathrm{Gal}(K/E) \in \mathcal{N}$. The automorphism τ fixes E since $\tau \in N$, so $\tau(a) = a$. Thus, $\tau = \mathrm{id}$, so $\bigcap_{N \in \mathcal{N}} N = \{\mathrm{id}\}$. For the second statement, if $\tau \in \sigma N$ for all N, then $\sigma^{-1}\tau \in N$ for all N; thus, $\sigma^{-1}\tau = \mathrm{id}$ by the first part. This yields $\tau = \sigma$, so $\bigcap_{N \in \mathcal{N}} \sigma N = \{\sigma\}$. □

Lemma 17.4 *Let $N_1, N_2 \in \mathcal{N}$. Then $N_1 \cap N_2 \in \mathcal{N}$.*

Proof. Let $N_i = \mathrm{Gal}(K/E_i)$ with $E_i \in \mathcal{I}$. Each E_i is finite Galois over F; hence, E_1E_2 is also finite Galois over F, so $E_1E_2 \in \mathcal{I}$. However,

$\text{Gal}(K/E_1E_2) = N_1 \cap N_2$; to see this, we note that $\sigma \in N_1 \cap N_2$ if and only if $\sigma|_{E_1} = \text{id}$ and $\sigma|_{E_2} = \text{id}$, if and only if $E_1 \subseteq \mathcal{F}(\sigma)$ and $E_2 \subseteq \mathcal{F}(\sigma)$, and if and only if $E_1E_2 \subseteq \mathcal{F}(\sigma)$. This last condition is true if and only if $\sigma \in \text{Gal}(K/E_1E_2)$. Thus, $N_1 \cap N_2 = \text{Gal}(K/E_1E_2) \in \mathcal{N}$. □

We can now define a topology on the Galois group G.

Definition 17.5 *The Krull topology on G is defined as follows: A subset X of G is open if $X = \varnothing$ or if $X = \bigcup_i \sigma_i N_i$ for some $\sigma_i \in G$ and $N_i \in \mathcal{N}$.*

From the definition, it is clear that G and $\varnothing$ are open sets and that the union of open sets is open. To show that we do indeed have a topology on G, it remains to see that the intersection of two open sets is again open. It is sufficient to show that $\tau_1 N_1 \cap \tau_2 N_2$ is open for any $N_1, N_2 \in \mathcal{N}$. To see this, if $\sigma \in \tau_1 N_1 \cap \tau_2 N_2$, then

$$\tau_1 N_1 \cap \tau_2 N_2 = \sigma N_1 \cap \sigma N_2 = \sigma(N_1 \cap N_2),$$

and $\sigma(N_1 \cap N_2)$ is open, since $N_1 \cap N_2 \in \mathcal{N}$ by Lemma 17.4.

We point out some properties of the Krull topology. Since each nonempty open set of G is a union of cosets of subgroups of $\mathcal{N}$, the set

$$\{\sigma N : \sigma \in G,\ N \in \mathcal{N}\}$$

is a basis for the Krull topology. If $N \in \mathcal{N}$, then $|G : N| < \infty$, so $G - \sigma N$ is a union of finitely many cosets of N. Therefore, σN is both open and closed. A set that is both closed and open is called *clopen*. The Krull topology thus has a basis of clopen sets. While the existence of nontrivial clopen sets is not common in more familiar topologies such as the usual topologies on $\mathbb{R}$ or $\mathbb{C}$, it is common for topologies arising in algebra. The following theorem describes the topology on G. Recall that a topological space is totally disconnected if the only connected subsets are single points.

Theorem 17.6 *As a topological space, G is Hausdorff, compact, and totally disconnected.*

Proof. If X is a subset of G and $\sigma, \tau \in X$, let σN be an open neighborhood of σ not containing τ. The existence of N follows from Lemma 17.3. Then

$$X = (\sigma N \cap X) \cup ((G - \sigma N) \cap X),$$

an intersection of two disjoint, nonempty open sets in X, so X is not connected. Therefore, G is totally disconnected. To show that G is Hausdorff, let $\sigma \in G$. Lemma 17.3 shows that $\{\sigma\} = \bigcap_N \sigma N$. If $\tau \neq \sigma$, then there is an $N \in \mathcal{N}$ with $\tau \notin \sigma N$. Each σN is an open neighborhood of σ but is also closed, as noted above. Thus, σN and $G - \sigma N$ are disjoint open sets with $\sigma \in \sigma N$ and $\tau \in G - \sigma N$, so G is Hausdorff.

The most difficult part of the proof is to show that G is compact. In proving that G is compact, we will indirectly show how G can be constructed from finite Galois groups. Let P be the direct product $\prod_{N \in \mathcal{N}} G/N$ of the finite groups G/N. We make P into a topological space by giving each G/N the discrete topology and then giving P the product topology. Note that each G/N is both Hausdorff and compact, so P is Hausdorff, and by the Tychonoff theorem, P is compact. There is a natural group homomorphism $f : G \longrightarrow P$ defined by $f(\sigma) = \{\sigma N\}$. We will show f is a homeomorphism from G to the image of f and that this image is a closed subset of P. Since P is compact and Hausdorff, this will show that $\mathrm{im}(f)$ is compact, hence G is compact, since G is homeomorphic to $\mathrm{im}(f)$.

Let f be as above. The kernel of f consists of those $\sigma \in G$ with $\{\sigma N\} = \{N\}$. Therefore, if $\sigma \in \ker(f)$, then $\sigma \in \bigcap_{N \in \mathcal{N}} N = \{\mathrm{id}\}$; this equality holds by Lemma 17.3. Thus, f is injective. Let $\pi_N : P \longrightarrow G/N$ be the projection onto the N-component. Then $\pi_N(f(\sigma)) = \sigma N$ for any $\sigma \in G$. The singleton sets τN form a basis for the discrete topology on G/N, so by definition of the product topology, every open set in P is a union of a finite intersection of sets of the form $\pi_N^{-1}(\tau N)$ for various $\tau \in G$ and $N \in \mathcal{N}$. To show that f is continuous, it is enough to show that $f^{-1}(\pi_N^{-1}(\{\tau N\}))$ is open in G for any τN. But this preimage is just τN, which is open, so f is continuous. Furthermore, $f(\tau N) = \pi_N^{-1}(\{\tau N\}) \cap \mathrm{im}(f)$ is open in $\mathrm{im}(f)$, so f^{-1} is also continuous. Therefore, f is a homeomorphism from G to $\mathrm{im}(f)$. It remains to show that $\mathrm{im}(f)$ is closed in P. In verifying that $\mathrm{im}(f)$ is closed in P, we will identify G/N with the isomorphic group $\mathrm{Gal}(E_N/F)$, where $E_N = \mathcal{F}(N)$. This isomorphism is from Lemma 17.2. This amounts to identifying the coset τN with $\tau|_{E_N}$. With this identification, for $\rho \in P$ the element $\pi_N(\rho)$ is an automorphism of E_N. Note that for $\tau \in G$ we have $\pi_N(f(\tau)) = \tau|_{E_N}$. Let

$$C = \{\, \rho \in P : \text{ for each } N, M \in \mathcal{N},\ \pi_N(\rho)|_{E_N \cap E_M} = \pi_M(\rho)|_{E_N \cap E_M} \,\}.$$

We claim that $C = \mathrm{im}(f)$. Now, $\mathrm{im}(f) \subseteq C$ since $\pi_N(f(\tau))|_{E_N} = \tau|_{E_N}$ for any $\tau \in G$. For the reverse inclusion, let $\rho \in C$. We define $\tau : K \longrightarrow K$ as follows. For $a \in K$, pick any $E_N \in \mathcal{I}$ with $a \in E_N$, possible by Lemma 17.1, and define $\tau(a) = \pi_N(\rho)(a)$. The condition on ρ to be an element of C shows that this is a well-defined map. To see that τ is a ring homomorphism, if $a, b \in K$, let $E_N \in \mathcal{I}$ with $a, b \in E_N$. Then $\tau|_{E_N} = \pi_N(\rho)$ is a ring homomorphism, so $\tau(a + b) = \tau(a) + \tau(b)$ and $\tau(ab) = \tau(a)\tau(b)$. The map τ is a bijection, since we can construct τ^{-1} by using ρ^{-1}. It is clear that τ fixes F, so $\tau \in G$. Now, as $\tau|_{E_N} = \pi_N(\rho)$, we see that $f(\tau) = \rho$. Thus, $C = \mathrm{im}(f)$. To show that C is closed in P, take any $\rho \in P$ with $\rho \notin C$. Then there are N, $M \in \mathcal{N}$ with $\pi_N(\rho)|_{E_N \cap E_M} \neq \pi_M(\rho)|_{E_N \cap E_M}$. Thus, $\pi_N^{-1}(\pi_N(\rho)) \cap \pi_M^{-1}(\pi_M(\rho))$ is an open subset of P containing ρ and disjoint from C. Therefore, $P - C$ is open, so $C = \mathrm{im}(f)$ is closed. □

The set $\mathcal{N}$, ordered by reverse inclusion, is a *directed set*; that is, if $N_1, N_2 \in \mathcal{N}$, then there is an $N_3 \in \mathcal{N}$ with $N_3 \subseteq N_1 \cap N_2$, namely $N_3 = N_1 \cap N_2$. The set $\{ G/N : N \in \mathcal{N} \}$ together with the natural projection maps $G/N_1 \longrightarrow G/N_2$ for $N_1 \subseteq N_2$ form a *directed system* of groups. The proof that $G = \mathrm{im}(f)$ can be viewed as showing that G is the *inverse limit* of the set of finite groups $\{G/N\}$ (see Problem 14). The inverse limit of a set of finite groups is called a *profinite group*. For more information on profinite groups, see Shatz [25], Serre [24], or Appendix C.

The next theorem is the final step we need to extend the fundamental theorem to arbitrary Galois extensions. This theorem shows how the topology on G comes in, and it is the analog of Proposition 2.14, which says that if G is a finite group of automorphisms of K, then $G = \mathrm{Gal}(K/\mathcal{F}(G))$.

Theorem 17.7 *Let H be a subgroup of G, and let $H' = \mathrm{Gal}(K/\mathcal{F}(H))$. Then $H' = \overline{H}$, the closure of H in the topology of G.*

Proof. It is clear that $H \subseteq H'$, so it suffices to show that H' is closed and that $H' \subseteq \overline{H}$. To show that H' is closed, take any $\sigma \in G - H'$. Then there is an $\alpha \in \mathcal{F}(H)$ with $\sigma(\alpha) \neq \alpha$. Take $E \in \mathcal{I}$ with $\alpha \in E$, and let $N = \mathrm{Gal}(K/E) \in \mathcal{N}$. Then, for any $\tau \in N$, we have $\tau(\alpha) = \alpha$, so $\sigma\tau(\alpha) = \sigma(\alpha) \neq \alpha$. Hence, σN is an open neighborhood of σ disjoint from H'. Therefore, $G - H'$ is open, so H' is closed. To prove the inclusion $H' \subseteq \overline{H}$, we first set $L = \mathcal{F}(H)$. Let $\sigma \in H'$ and $N \in \mathcal{N}$. Set $E = \mathcal{F}(N) \in \mathcal{I}$, and let $H_0 = \{\rho|_E : \rho \in H\}$, a subgroup of the finite group $\mathrm{Gal}(E/F)$. Since $\mathcal{F}(H_0) = \mathcal{F}(H) \cap E = L \cap E$, the fundamental theorem for finite Galois extensions shows that $H_0 = \mathrm{Gal}(E/(E \cap L))$. Since $\sigma \in H'$, we have $\sigma|_L = \mathrm{id}$, so $\sigma|_E \in H_0$. Therefore, there is a $\rho \in H$ with $\rho|_E = \sigma|_E$. Thus, $\sigma^{-1}\rho \in \mathrm{Gal}(K/E) = N$, so $\rho \in \sigma N \cap H$. This shows that every basic open neighborhood σN of $\sigma \in H'$ meets H, so $\sigma \in \overline{H}$. This proves the inclusion $H' \subseteq \overline{H}$ and finishes the proof. □

A way to describe $H' = \mathrm{Gal}(K/\mathcal{F}(H))$ that does not involve the topology on G is $H' = \bigcap_{N \in \mathcal{N}} HN$ (see Problem 1).

Theorem 17.8 (Fundamental Theorem of Infinite Galois Theory) *Let K be a Galois extension of F, and let $G = \mathrm{Gal}(K/F)$. With the Krull topology on G, the maps $L \mapsto \mathrm{Gal}(K/L)$ and $H \mapsto \mathcal{F}(H)$ give an inclusion reversing correspondence between the fields L with $F \subseteq L \subseteq K$ and the closed subgroups H of G. Furthermore, if $L \leftrightarrow H$, then $|G : H| < \infty$ if and only if $[L : F] < \infty$, if and only if H is open. When this occurs, $|G : H| = [L : F]$. Also, H is normal in G if and only if L is Galois over F, and when this occurs, there is a group isomorphism $\mathrm{Gal}(L/F) \cong G/N$. If G/N is given the quotient topology, this isomorphism is also a homeomorphism.*

Proof. If L is a subfield of K containing F, then K is normal and separable over L, so K is Galois over L. Thus, $L = \mathcal{F}(\operatorname{Gal}(K/L))$. If H is a subgroup of G, then Theorem 17.7 shows that $H = \operatorname{Gal}(K/\mathcal{F}(H))$ if and only if H is closed. The two maps $L \mapsto \operatorname{Gal}(K/L)$ and $H \mapsto \mathcal{F}(H)$ then give an inclusion reversing correspondence between the set of intermediate fields of K/F and the set of closed subgroups of G.

Let L be an intermediate field of K/F, and let $H = \operatorname{Gal}(K/L)$. Suppose that $|G : H| < \infty$. Then $G - H$ is a finite union of cosets of H, each of which is closed, since H is closed. Thus, $G - H$ is closed, so H is open. Conversely, if H is open, then H contains some basic neighborhood of id, so $N \subseteq H$ for some $N \in \mathcal{N}$. If $E = \mathcal{F}(N)$, then $L \subseteq E$, so $[L : F] < \infty$. Finally, if $[L : F] < \infty$, then choose an $E \in \mathcal{I}$ with $L \subseteq E$, possible by Lemma 17.1. Let $N = \operatorname{Gal}(K/E)$. Then $N \subseteq H$, since $L \subseteq E$, so $|G : H| \leq |G : N| < \infty$. By Lemma 17.2, we have $G/N \cong \operatorname{Gal}(E/F)$ via the map $\sigma N \mapsto \sigma|_E$. Thus, H/N maps to $\{\rho|_E : \rho \in H\}$, a subgroup of $\operatorname{Gal}(E/F)$ with fixed field $L \cap E = L$. By the fundamental theorem for finite extensions, the order of this group is $[E : L]$. Therefore,

$$\begin{aligned} |G : H| = |G/N : H/N| &= \frac{|G/N|}{|H/N|} = \frac{[E : F]}{[E : L]} \\ &= [L : F]. \end{aligned}$$

For the statement about normality, we continue to assume that $H = \operatorname{Gal}(K/L)$. Suppose that H is a normal subgroup of G. Let $a \in L$, and let $f(x) = \min(F, a)$. If $b \in K$ is any root of f, by the isomorphism extension theorem there is a $\sigma \in G$ with $\sigma(a) = b$. To see that $b \in L$, take $\tau \in H$. Then

$$\begin{aligned} \tau(b) &= \sigma^{-1}(\sigma\tau\sigma^{-1}(a)) \\ &= \sigma^{-1}(a) = b \end{aligned}$$

since $\sigma\tau\sigma^{-1} \in H$, as H is normal in G. Thus, $b \in \mathcal{F}(H) = L$, so f splits over L. This proves that L is normal over F, and L is separable over F since K/F is separable. Therefore, L is Galois over F. Conversely, if L is Galois over F, then by the remark before Lemma 17.1 we see that the map $\sigma \mapsto \sigma|_L$ is a well-defined group homomorphism $\theta : G \longrightarrow \operatorname{Gal}(L/F)$. The kernel of θ is $\operatorname{Gal}(K/L) = H$, so H is normal in G, and θ is surjective by an application of the isomorphism extension theorem. Thus, $G/H \cong \operatorname{Gal}(L/F)$.

The last step of the proof is to show that the natural map $\nu : G/H \longrightarrow \operatorname{Gal}(L/F)$ is a homeomorphism when H is normal in G. Note that a basic open subset of $\operatorname{Gal}(L/F)$ has the form $\rho \operatorname{Gal}(L/E)$ for some extension E that is finite Galois over F and is contained in L. Let $N = \operatorname{Gal}(K/E) \in \mathcal{N}$. Then $\theta^{-1}(\operatorname{Gal}(L/E)) = N$. Thus, $\theta^{-1}(\rho \operatorname{Gal}(L/E)) = \tau N$ for any $\tau \in G$ with $\tau|_L = \rho$, so this preimage is open in G. Therefore, θ is continuous. Furthermore, the image of a compact set under a continuous map is compact, and any compact subset of a Hausdorff space is closed. Since G is

compact and $\mathrm{Gal}(L/F)$ is Hausdorff, θ maps closed sets to closed sets; that is, θ is a closed map. The map $\nu : G/H \longrightarrow \mathrm{Gal}(L/F)$ induced from θ is then also continuous and closed when G/H is given the quotient topology, so ν is a homeomorphism. □

Example 17.9 Let K/F be a Galois extension with $[K : F] < \infty$. Then the Krull topology on $\mathrm{Gal}(K/F)$ is the discrete topology; hence, every subgroup of $\mathrm{Gal}(K/F)$ is closed. Thus, we recover the original fundamental theorem of Galois theory from Theorem 17.8.

Example 17.10 Let $K = \mathbb{Q}(\{e^{2\pi ik/n} : k, n \in \mathbb{N}\})$ be the field generated over $\mathbb{Q}$ by all roots of unity in $\mathbb{C}$. Then K is the splitting field over $\mathbb{Q}$ of the set $\{x^n - 1 : n \in \mathbb{N}\}$, so $K/\mathbb{Q}$ is Galois. If L is a finite Galois extension of $\mathbb{Q}$ contained in K, then L is contained in a cyclotomic extension of $\mathbb{Q}$. The Galois group of a cyclotomic extension is Abelian. Consequently, $\mathrm{Gal}(L/F)$ is Abelian. To see that $\mathrm{Gal}(K/F)$ is Abelian, by the proof of Theorem 17.8 the Galois group $\mathrm{Gal}(K/F)$ is isomorphic to a subgroup of the direct product of the $\mathrm{Gal}(L/F)$ as L ranges over finite Galois subextensions of $\mathbb{Q}$, so $\mathrm{Gal}(K/F)$ is Abelian. As a consequence of this fact, any subextension of $K/\mathbb{Q}$ is a Galois extension of $\mathbb{Q}$.

We give an alternate proof that $\mathrm{Gal}(K/F)$ is Abelian that does not use the proof of Theorem 17.8. Take $\sigma, \tau \in \mathrm{Gal}(K/\mathbb{Q})$. If $a \in K$, then there is an intermediate field L of $K/\mathbb{Q}$ that is Galois over $\mathbb{Q}$ and that $a \in L$. The restrictions $\sigma|_L, \tau|_L$ are elements of $\mathrm{Gal}(L/\mathbb{Q})$, and this group is Abelian by the previous paragraph. Thus,

$$\sigma(\tau(a)) = \sigma|_L(\tau|_L(a)) = \tau|_L(\sigma|_L(a)) = \tau(\sigma(a)).$$

Consequently, $\sigma\tau = \tau\sigma$, so $\mathrm{Gal}(K/\mathbb{Q})$ is Abelian.

Example 17.11 Let K be an algebraic closure of $\mathbb{F}_p$. Since $\mathbb{F}_p$ is perfect, K is separable, and hence K is Galois over $\mathbb{F}_p$. Let $\sigma : K \longrightarrow K$ be defined by $\sigma(a) = a^p$. Then $\sigma \in G = \mathrm{Gal}(K/\mathbb{F}_p)$, and the fixed field of the cyclic subgroup H of G generated by σ is $\mathbb{F}_p$. However, we prove that $H \neq G$ by constructing an automorphism of K that is not in H. To see this, pick an integer n_r for each $r \in \mathbb{N}$ such that if r divides s, then $n_s \equiv n_r \pmod{r}$. If F_r is the subfield of K containing p^r elements, then define τ by $\tau(a) = \sigma^{n_r}(a)$ if $a \in F_r$. The conditions on the n_r show that τ is well defined, and an easy argument shows that τ is an automorphism of K that fixes $\mathbb{F}_p$. For a specific example of a choice of the n_r, for $r \in \mathbb{N}$, write $r = p^m q$ with q not a multiple of p. Let n_r satisfy

$$\begin{aligned} n_r &\equiv 1 + p + \cdots + p^{m-1} \pmod{p^m}, \\ n_r &\equiv 0 \pmod{q}. \end{aligned}$$

Such integers exist by the Chinese remainder theorem of number theory, since p^m and q are relatively prime. If $\tau = \sigma^t$ for some t, then for all

r, $\tau|_{F_r} = \sigma^t|_{F_r}$, so $n_r \equiv t \pmod{r}$, as $\mathrm{Gal}(F_r/\mathbb{F}_p)$ is the cyclic group generated by $\sigma|_{F_r}$, which has order r. This cannot happen as $n_{p^m} \longrightarrow \infty$ as $m \longrightarrow \infty$. Therefore, $\tau \notin H$, so H is not a closed subgroup of G. The group G is obtained topologically from H, since $G = \overline{H}$ by Theorem 17.7. The argument that $G = \mathrm{im}(f)$ in the proof of Theorem 17.6 shows that any element of G is obtained by the construction above, for an appropriate choice of the n_r. This gives a description of the Galois group G as

$$\mathrm{Gal}(K/\mathbb{F}_p) \cong \left\{ \{n_r\} \in \prod_r \mathbb{F}_{p^r} \ : \text{if } r \text{ divides } s, \text{ then } n_s \equiv n_r \pmod{r} \right\}.$$

Problems

Unless otherwise stated, in the following problems K/F will be an infinite Galois extension with $G = \mathrm{Gal}(K/F)$.

1. Let H be a subgroup of $\mathrm{Gal}(K/F)$. Show that the closure $\overline{H}$ of H with respect to the Krull topology on $\mathrm{Gal}(K/F)$ is $\overline{H} = \bigcap_{N \in \mathcal{N}} HN$.

2. Let L an intermediate field of K/F. Show that the Krull topology on $\mathrm{Gal}(K/L)$ is the subspace topology inherited from the Krull topology on $\mathrm{Gal}(K/F)$.

3. Show that $\mathrm{Gal}(K/F)$ is uncountable. Use this to give an example of a Galois extension K/F with $[K : F] \neq |\mathrm{Gal}(K/F)|$.
(Hint: Obtain a chain of finite degree Galois extensions of F whose union is K, and use the isomorphism extension theorem.)

4. Show that there are subgroups of $\mathrm{Gal}(K/F)$ that are not closed.

5. Here is an alternative, purely topological way to prove Problem 3. Prove that a totally disconnected compact topological space X with no isolated points is uncountable, provided that $|X| = \infty$.

6. Let H be a subgroup of $\mathrm{Gal}(K/F)$.

   (a) If H is open in the Krull topology, show that H has finite index in $\mathrm{Gal}(K/F)$.

   (b) If H has finite index in $\mathrm{Gal}(K/F)$, show that H is open if and only if H is closed.

7. Give an example of an extension K/F such that $\mathrm{Gal}(K/F)$ contains a subgroup of finite index that is neither open nor closed.

8. Let K/F be an infinite Galois extension, and let N be a normal subgroup of $\mathrm{Gal}(K/F)$. Show that $\overline{N}$ is a normal subgroup of $\mathrm{Gal}(K/F)$.

9. Let K/F be a Galois extension, and let H be a subgroup of $\mathrm{Gal}(K/F)$. Show that H is dense in $\mathrm{Gal}(K/F)$ if and only if for every finite normal intermediate field L, every F-automorphism of L is the restriction to L of some element of H.

10. Use the previous problem to show that H is dense in $\mathrm{Gal}(K/F)$ if and only if for each finite Galois intermediate field L, we have

$$\mathrm{Gal}(L/F) \cong H/(H \cap \mathrm{Gal}(K/L)).$$

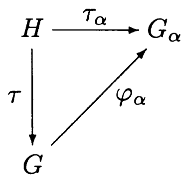
11. Let K be a Galois extension of F, and let $G = \mathrm{Gal}(K/F)$. Show that the multiplication map $G \times G$ to G given by $(g, h) \mapsto gh$ is continuous with respect to the Krull topology, as is the inverse map $\sigma \mapsto \sigma^{-1}$. This means that G is a *topological group*.

12. Here is an alternative way to view the Krull topology on a Galois group. Let K/F be a Galois extension. Let K have the discrete topology, and let K^K have the product topology. The Galois group $\mathrm{Gal}(K/F)$ is a subset of K^K. Show that $\mathrm{Gal}(K/F)$ is a closed subset of K^K, and notice that the same argument shows that $\mathrm{Gal}(K/L)$ is also closed if L is an intermediate field of K/F. Moreover, show that the Krull topology on $\mathrm{Gal}(K/F)$ is the same as the subspace topology.
(Note: This topology on K^K is called the finite topology, and it is the same as the compact open topology on K^K.)

13. This problem describes inverse limits. Problem 14 shows that a Galois group of a Galois extension is an inverse limit of finite Galois groups. Let $\{G_\alpha\}_{\alpha \in I}$ be a set of groups. Suppose that I is a *directed set*; that is, I has a partial order $\leq$, such that for any $\alpha, \beta \in I$, there is a $\gamma \in I$ with $\alpha \leq \gamma$ and $\beta \leq \gamma$. Assume that for each pair $\alpha \leq \beta$ in I there is a group homomorphism $\varphi_{\alpha\beta} : G_\beta \to G_\alpha$ satisfying the following compatibility conditions:

   - $\varphi_{\alpha\alpha} = \mathrm{id}$;
   - if $\alpha \leq \beta \leq \gamma$, then $\varphi_{\alpha\gamma} = \varphi_{\beta\gamma} \circ \varphi_{\alpha\beta}$.

   A group G together with homomorphisms $\varphi_\alpha : G \to G_\alpha$ satisfying $\varphi_{\alpha\beta} \circ \varphi_\beta = \varphi_\alpha$ for each pair $\alpha \leq \beta$ is said to be an *inverse limit* of the G_α (along with the maps $\varphi_{\alpha\beta}$), provided that G satisfies the following universal mapping property: If H is a group together with homomorphisms $\tau_\alpha : H \to G_\alpha$ satisfying $\varphi_{\alpha\beta} \circ \tau_\beta = \tau_\alpha$ for each pair $\alpha \leq \beta$, then there is a unique homomorphism $\tau : H \to G$ with $\tau_\alpha = \varphi_\alpha \circ \tau$ for each α; that is, the following diagram commutes:



(a) Show that any two inverse limits of $\{G_\alpha\}$ are isomorphic.

(b) Show that inverse limits exist: Let

$$G = \left\{ \{g_\alpha\} \in \prod_{\alpha \in I} G_\alpha : g_\alpha = \varphi_{\alpha\beta}(g_\beta) \text{ for all } \alpha, \beta \text{ with } \alpha \leq \beta \right\}.$$

Show that G is an inverse limit of the $\{G_\alpha\}$, where the maps φ_α are the restrictions to G of the usual projection maps.

14. Let K/F be a Galois extension, and let $G = \text{Gal}(K/F)$. Let $\mathcal{N}$ be as in the section, and order $\mathcal{N}$ by reverse inclusion. Let $\varphi_{\alpha\beta} : G/N_\alpha \to G/N_\beta$ be the canonical projection whenever $N_\alpha \subseteq N_\beta$. Show that $\{G/N : N \in \mathcal{N}\}$ is an inverse system of groups and that G is the inverse limit of this system.

15. Let G be a group.

    (a) Show that the set $\mathcal{S}$ of normal subgroups of G of finite index, ordered by inclusion, is a directed set.

    (b) Let $\varphi_{\alpha\beta} : G/N_\alpha \to G/N_\beta$ be the natural projection map when $N_\alpha \subseteq N_\beta$, and let $\widehat{G}$ be the inverse limit of the groups $\{G/N_\alpha\}_{\alpha \in \mathcal{S}}$. Show that there is a natural homomorphism $\varphi : G \to \widehat{G}$ and that φ is injective if the intersection of all the N_α is $\langle e \rangle$.

    (c) Let G be a profinite group. Show that $G \cong \widehat{G}$.

16. If K is the algebraic closure of $\mathbb{F}_p$, show that $\text{Gal}(K/\mathbb{F}_p) \cong \widehat{\mathbb{Z}}$.

17. Let G be a profinite group. Show that $G = \text{Gal}(K/F)$ for some Galois extension K/F.

18 Some Infinite Galois Extensions

In this section, we describe some examples of infinite Galois extensions. Some of these extensions will arise from group theoretic properties of infinite Galois groups. To discuss some of these extensions, we will require knowledge of profinite groups, information about which can be found in Appendix C, Shatz [25], or Serre [24].

The separable closure of a field

Let F be a field. Then F is said to be *separably closed* if there is no proper separable extension of F. Let F_{ac} be an algebraic closure of F. Then F_{ac} is the splitting field of the set of all nonconstant polynomials in $F[x]$; hence, F_{ac} is a normal extension of F. However, if F is not perfect, then F_{ac} is not Galois over F. Let F_s be the separable closure of F in F_{ac}. The field F_s is called the *separable closure* of F. The following description of F_s follows quickly from the properties of normal extensions.

Proposition 18.1 *Let F_s be the separable closure of the field F. Then F_s is Galois over F with $\mathrm{Gal}(F_s/F) \cong \mathrm{Gal}(F_{ac}/F)$. Moreover, F_s is a maximal separable extension of F, meaning that F_s is not properly contained in any separable extension of F. Thus, F_s is separably closed.*

Proof. The field F_s is Galois over F, and $\mathrm{Gal}(F_s/F) = \mathrm{Gal}(F_{ac}/F)$ by Theorem 4.23. Suppose that $F_s \subseteq L$ with L/F separable. Then we can embed $L \subseteq F_{ac}$, and then $L = F_s$, since F_s is the set of all separable elements over F in F_{ac}. Finally, if L is a separable extension of F_s, then by transitivity of separability, L is a separable extension of F, so $L = F_s$. Therefore, F_s is separably closed. □

The group $\mathrm{Gal}(F_s/F) \cong \mathrm{Gal}(F_{ac}/F)$ is often called the *absolute Galois group* of F. If G is the Galois group of a Galois extension of F, then G is a homomorphic image of $\mathrm{Gal}(F_s/F)$ by the fundamental theorem.

The quadratic closure of a field

In the next three sections, we require some knowledge of profinite groups. If G is a profinite group and p is a prime, then G is a pro-p-group if every open normal subgroup of G has index in G equal to a power of p. If $G = \mathrm{Gal}(K/F)$ for a Galois extension K/F, then G is a pro-p-group if and only if every finite Galois subextension of K/F has degree a power of p over F.

Let F be a field of characteristic not 2. Then F is said to be *quadratically closed* if there is no proper quadratic extension of F. The *quadratic closure* F_q of F is a subfield of F_s that is quadratically closed and is a Galois extension of F with $\mathrm{Gal}(F_q/F)$ a pro-2-group. The following proposition shows the existence and uniqueness of the quadratic closure of a field.

Proposition 18.2 *Let F be a field with $\mathrm{char}(F) \neq 2$. Then the quadratic closure F_q of F is the composite inside a fixed algebraic closure of F of all Galois extensions of F of degree a power of 2.*

Proof. Let K be the composite inside a fixed algebraic closure of F of all Galois extensions of F of degree a power of 2. Then K is Galois over F. To show that $G = \mathrm{Gal}(K/F)$ is a pro-2-group, let N be an open normal

subgroup of G. If $L = \mathcal{F}(H)$, then $[L : F] = [G : N]$ by the fundamental theorem. The intermediate field L is a finite extension of F; hence, L lies in a composite of finitely many Galois extensions of F of degree a power of 2. Any such composite has degree over F a power of 2 by the theorem of natural irrationalities, so $[L : F]$ is a power of 2. Thus, $[G : N]$ is a power of 2, so G is a pro-2-group.

To see that K is quadratically closed, suppose that L/K is a quadratic extension, and say $L = K(\sqrt{a})$ for some $a \in K$. Then $a \in E$ for some finite Galois subextension E. By the argument above, we have $[E : F] = 2^r$ for some r. The extension $E(\sqrt{a})/E$ has degree at most 2. If $\sqrt{a} \in E$, then $L = K$ and we are done. If not, consider the polynomial

$$\prod_{\sigma \in \text{Gal}(E/F)} (x^2 - \sigma(a)) \in F[x].$$

The splitting field N over F of this polynomial is $N = F(\{\sqrt{\sigma(a)} : \sigma \in \text{Gal}(E/F)\})$. Hence, N is a 2-Kummer extension of F, so $[N : F]$ is a power of 2. The field N is a Galois extension of F of degree a power of 2, so $N \subseteq K$. Moreover, $\sqrt{a} \in N$. This shows that $\sqrt{a} \in K$, so $L = K$. Thus, K is quadratically closed. □

In the next proposition, we give an alternate description of the quadratic closure of a field F of characteristic not 2.

Proposition 18.3 *Let F be a field of characteristic with* $\text{char}(F) \neq 2$. *We define fields $\{F_n\}$ by recursion by setting $F_0 = F$ and $F_{n+1} = F_n(\{\sqrt{a} : a \in F_n\})$. Then the quadratic closure of F is the union $\bigcup_{n=1}^{\infty} F_n$.*

Proof. Let $K = \bigcup_{n=1}^{\infty} F_n$. Then K is a field, since $\{F_n\}$ is a totally ordered collection of fields. We show that K is quadratically closed. If $a \in K$, then $a \in F_n$ for some n, so $\sqrt{a} \in F_{n+1} \subseteq K$. Thus, $K(\sqrt{a}) = K$, so K is indeed quadratically closed. Let F_q be the quadratic closure of F. Then $\sqrt{a} \in F_q$ for each $a \in F_q$, so we see that $F_1 \subseteq F_q$. Suppose that $F_n \subseteq F_q$. The reasoning we used to show that K is quadratically closed shows also that $F_{n+1} \subseteq F_q$, so $K \subseteq F_q$. To see that this inclusion is an equality, let E be a Galois extension of F of degree a power of 2. Then EK/K has degree a power of 2 by natural irrationalities. If $[EK : K] > 1$, then the group $\text{Gal}(EK/K)$ has a subgroup of index 2 by the theory of p-groups. If L is the fixed field of this subgroup, then $[L : K] = 2$. However, this is impossible, since K is quadratically closed. This forces $EK = K$, so $E \subseteq K$. Since F_q is the composite of all such E, we see that $F_q \subseteq K$, so $K = F_q$. □

The p-closure of a field

Let F be a field of characteristic not p, where p is some prime. Fix some algebraic closure F_{ac} of F. The *p-closure* F_p of F is the composite in F_{ac} of

all Galois extensions of F of degree a power of p. The quadratic closure of F is then just F_2. The basic properties of the p-closure of a field are given in the following results. The first describes what finite extensions of F lie inside F_p.

Lemma 18.4 *Let p be a prime, and let F be a field with* $\text{char}(F) \neq p$. *If L is an intermediate field of F_{ac}/F with $[L : F] < \infty$, then $L \subseteq F_p$ if and only if L lies in a Galois extension of F of degree a power of p. In particular, any finite intermediate field of F_p/F has degree over F a power of p.*

Proof. If L is a field lying inside some Galois extension E of F with $[E : F]$ a power of p, then $E \subseteq F_p$, so $L \subseteq F_p$. Conversely, suppose that $L \subseteq F_p$ and $[L : F] < \infty$. Then $L = F(a_1, \ldots, a_n)$ for some $a_i \in L$. From the definition of F_p, for each i there is a Galois extension E_i/F such that $a_i \in E_i$ and $[E_i : F]$ is a power of p. The composition of the E_i is a Galois extension of F, whose degree over F is also a power of p by natural irrationalities. □

Proposition 18.5 *Let F_p be the p-closure of a field F with* $\text{char}(F) \neq p$. *Then F_p is a Galois extension of F and* $\text{Gal}(F_p/F)$ *is a pro-p-group. Moreover, F_p has no Galois extensions of degree p.*

Proof. The proof that F_p/F is Galois with $\text{Gal}(F_p/F)$ a pro-p-group is essentially the same as the proof for the corresponding result about the quadratic closure, so we do not repeat it here. For the final statement, suppose that L is a Galois extension of F_p with $[L : F_p] = p$. We need to obtain a contradiction. The argument we gave for the corresponding result about the quadratic closure will not work, since the composite of field extensions of degree a power of p need not have degree a power of p if $p \neq 2$. Instead, we argue as follows. Say $L = F_p(a)$, and let $a_1, a_2, \ldots, a_p$ be the roots of $\min(F_p, a)$. Since $F_p(a)/F_p$ is Galois, each $a_i \in F_p(a)$. By the construction of F_p, for each i we can find a finite Galois extension E_i/F of degree a power of p with $a_i \in E_i(a)$ and $\min(F_p, a) \in E_i$. Taking the composite of all the E_i, we obtain a finite Galois extension E/F of degree a power of p with $a_i \in E(a)$ and $\min(F_p, a) \in E$. Therefore, $E(a)/E$ is Galois of degree p.

Let $f(x) = \prod_{\sigma \in \text{Gal}(E/F)} \sigma(g(x))$, a polynomial over F with $f(a) = 0$. For each σ, let a_σ be a root of $\sigma(g)$. Let N be the normal closure of $F(a)/F$, so N is the splitting field of $f(x)$ over F. The field NE is normal over F; hence, by the isomorphism extension theorem, for each $\sigma \in \text{Gal}(E/F)$ there is a $\sigma' \in \text{Gal}(NE/F)$ extending σ with $\sigma(a) = a_\sigma$. The automorphism σ' sends $E(a)$ to $E(a_\sigma)$. Since all the roots of g lie in $E(a)$, all the roots of $\sigma(g)$ lie in $E(a_\sigma)$. Thus, for each σ, the extension $E(a_\sigma)/E$ is Galois and is of degree p. However, $NE = E(\{a_\sigma\})$, so NE is a composite over E of Galois extensions of degree p; hence, $[NE : E]$ is a power of p by natural irrationalities. Therefore, $[NE : F]$ is a power of p, so $a \in F(a) \subseteq NE$

forces $a \in F_p$. This is a contradiction, so F_p has no Galois extension of degree p. □

If F contains a primitive pth root of unity, then there is a construction of F_p analogous to that of the quadratic closure of F.

Proposition 18.6 *Suppose that F contains a primitive pth root of unity. Define a sequence of fields $\{F_n\}$ by recursion by setting $F_0 = F$ and $F_{n+1} = F_n(\{\sqrt[p]{a} : a \in F_n\})$. Then the p-closure of F is $\bigcup_{n=1}^{\infty} F_n$.*

Proof. The proof is essentially the same as that for the quadratic closure, so we only outline the proof. If $F_n \subseteq F_p$ and $a \in F_n$, then either $F_n(\sqrt[p]{a}) = F_n$, or $F_n(\sqrt[p]{a})/F_n$ is a Galois extension of degree p, by Proposition 9.6. In either case, $F_n(\sqrt[p]{a}) \subseteq F_p$ by the previous proposition. This shows that $\bigcup_{n=1}^{\infty} F_n \subseteq F_p$. To get the reverse inclusion, let E/F be a Galois extension of degree a power of p. By the theory of p-groups and the fundamental theorem of Galois theory, there is a chain of intermediate fields

$$F = E_0 \subseteq E_1 \subseteq \cdots \subseteq E_n = E$$

with E_{i+1}/E_i Galois of degree p. Since F contains a primitive pth root of unity, $E_{i+1} = E_i(\sqrt[p]{a_i})$ for some $a_i \in E_i$ by Theorem 9.5. By induction, we can see that $E_i \subseteq F_i$, so $E \subseteq \bigcup_{n=1}^{\infty} F_n$. Since F_p is the composite of all such E, this gives the reverse inclusion we want. □

The maximal prime to p extension

Let G be a profinite group, and suppose that p divides $|G|$. Then a p-Sylow subgroup of G is a pro-p-group H such that $[G : H]$ is prime to p. Recall that a profinite group has a p-Sylow subgroup for every prime divisor p of $|G|$ and that any two p-Sylow subgroups of G are conjugate.

Let F be a perfect field and let p be a prime. If $G = \mathrm{Gal}(F_s/F)$, let P be a p-Sylow subgroup of G. If K is the fixed field of P, then K is called the *maximal prime to p extension* of F. The maximal prime to p extension of a field is not uniquely determined. However, since any two p-Sylow subgroups of a profinite group are conjugate, any two maximal prime to p extensions of F are F-isomorphic. The reason for the terminology above can be found in the following result.

Proposition 18.7 *Let F be a field, let p be a prime, and let K be a maximal prime to p extension of F. Then any finite extension of K has degree a power of p, and if L is an intermediate field of K/F with $[L : F] < \infty$, then $[L : F]$ is relatively prime to p. Moreover, any separable field extension L of F with $[L : F]$ relatively prime to p is contained in some maximal prime to p extension of F.*

Proof. Recall that if U is an open subgroup of a p-Sylow subgroup P of $G = \mathrm{Gal}(F_s/F)$, then $[P : U]$ is a power of p, and if V is open in G with $P \subseteq V \subseteq G$, then $[G : V]$ is relatively prime to p. Suppose that M is a finite extension of K. If $H = \mathrm{Gal}(F_s/M)$, then by the fundamental theorem, we have $[P : H] = [M : K] < \infty$, so H is an open subgroup of P. Thus, $[P : H]$ is a power of p, so $[M : K]$ is a power of p.

For the second statement, let L be an intermediate field of K/F with $[L : F] < \infty$. If $A = \mathrm{Gal}(F_s/L)$, then $P \subseteq A$ and $[G : A] = [L : F]$ is finite, by the fundamental theorem. Since $[G : A]$ is relatively prime to p, we see that $[L : F]$ is relatively prime to p.

Let L/F be an extension with $[L : F]$ relatively prime to p. Let F_s be the separable closure of F, and let $G = \mathrm{Gal}(F_s/F)$. Set $H = \mathrm{Gal}(F_s/L)$, a closed subgroup of G, and let P' be a p-Sylow subgroup of H. There is a p-Sylow subgroup P of G that contains P'. Note that $[G : H] = [L : F]$ is relatively prime to p. Moreover, we have

$$\begin{aligned}[G : P'] &= [G : H] \cdot [H : P'] \\ &= [G : P] \cdot [P : P'].\end{aligned}$$

Both $[G : H]$ and $[H : P']$ are supernatural numbers not divisible by p, so $[P : P']$ is not divisible by p. But, since P is a pro-p-group, $[P : P']$ is a power of p. This forces $[P : P'] = 1$, so $P' = P$. Therefore, $P \subseteq H$, and so $L = \mathcal{F}(H)$ is contained in $\mathcal{F}(P)$, a maximal prime to p extension of F. □

Example 18.8 The maximal prime to p extension of a field F need not be the composite of all finite extensions of degree relatively prime to p. For example, if $F = \mathbb{Q}$ and $p = 3$, then $\mathbb{Q}(\sqrt[3]{5})$ and $\mathbb{Q}(\omega\sqrt[3]{5})$ are both of degree 3 over $\mathbb{Q}$, where ω is a primitive third root of unity, but their composite is $\mathbb{Q}(\omega, \sqrt[3]{5})$, which has degree 6 over $\mathbb{Q}$. Therefore, these fields are not both contained in a common maximal prime to p extension of $\mathbb{Q}$.

Problem 5 addresses the construction of a maximal prime to p extension when F is not perfect.

The maximal Abelian extension

Let F be a field, and let F_s be the separable closure of F. Let $G = \mathrm{Gal}(F_s/F)$. If G' is the commutator subgroup of G, then the fixed field F_a of G' is called the *maximal Abelian extension* of F. This name is justified by the following result.

Proposition 18.9 *Let F_a be the maximal Abelian extension of a field F. Then F_a/F is a Galois extension and* $\mathrm{Gal}(F_a/F)$ *is an Abelian group. The field F_a has no extensions that are Abelian Galois extensions of F. Moreover, F_a is the composite in F_s of all finite Abelian Galois extensions of F.*

Proof. The commutator subgroup G' of G is a normal subgroup, so the closure $\overline{G'}$ of G' is a closed normal subgroup of G (see Problem 17.8). Thus, by the fundamental theorem, $F_a = \mathcal{F}(\overline{G'})$ is a Galois extension of F and $\text{Gal}(F_a/F) \cong G/\overline{G'}$. The group $G/\overline{G'}$ is a homomorphic image of the Abelian group G/G', so $G/\overline{G'}$ is also Abelian.

If $L \supseteq F_a$ is an Abelian Galois extension of F, then $L \subseteq F_s$. Let $H = \text{Gal}(F_s/L)$, a subgroup of G'. However, $G/H \cong \text{Gal}(L/F)$, so G/H is Abelian. Thus, $G' \subseteq H$, so $H = G'$. Therefore, F_a is not properly contained in any Abelian extension of F.

For the final statement, if K/F is finite Abelian Galois, then KF_a/F_a is Abelian Galois by natural irrationalities. Thus, $KF_a = F_a$, so $K \subseteq F_a$. Since every element of F_a lies in a finite Galois extension of F, to show that F_a is the composite of all finite Abelian Galois extensions of F it suffices to show that every finite Galois extension of F inside F_a is an Abelian extension. Let E be such an extension. If $H = \text{Gal}(F_s/E)$, then H is a normal subgroup of G containing G'; hence, G/H is Abelian. But, by the fundamental theorem, we have $\text{Gal}(E/F) \cong G/H$, so E/F is an Abelian Galois extension. □

Example 18.10 The Kronecker–Weber theorem of algebraic number theory states that any Abelian extension of $\mathbb{Q}$ is contained in a cyclotomic extension. Consequently, the maximal Abelian extension of $\mathbb{Q}$ is the infinite cyclotomic extension $\mathbb{Q}(\{\omega_n : n \in \mathbb{N}\})$.

Example 18.11 If F is a field containing a primitive nth root of unity for all n, then the maximal Abelian extension of F is $F(\{\sqrt[n]{a} : a \in F, n \in \mathbb{N}\})$. This follows from Kummer theory (see Problem 11.6 for part of this claim).

Problems

1. Let L be an intermediate field of the p-closure F_p of a field F. If $[L : F] < \infty$, show that there is a chain of fields

$$F = L_0 \subseteq L_1 \subseteq \cdots \subseteq L_n = L$$

such that for each i the extension L_{i+1}/L_i is Galois of degree p.

2. Show that the p-closure of a field has no Galois extensions of degree p^n for any $n > 1$.

3. Describe the p-closure of $\mathbb{F}_p$.

4. Describe the maximal prime to p extension of $\mathbb{F}_p$.

5. Let F be a field, not necessarily perfect. Let F_{ac} and F_s be the algebraic and separable closures of F, respectively, and let I be the purely

inseparable closure of F in F_{ac}. If p is a prime, let P be a p-Sylow subgroup of $\mathrm{Gal}(F_s/F)$, and let $K = \mathcal{F}(P) \subseteq F_s$. If $p \neq \mathrm{char}(F)$, call KI a maximal prime to p extension of F, and if $\mathrm{char}(F) = p$, call I a maximal prime to p extension of F. Prove the analog of Proposition 18.7 in this case.

V

Transcendental Extensions

In this chapter, we study field extensions that are not algebraic. In the first two sections, we give the main properties of these extensions. In the remaining sections, we focus on finitely generated extensions. We discuss how these extensions arise in algebraic geometry and how their study can lead to geometric information, and we use algebraic analogs of derivations and differentials to study these extensions.

19 Transcendence Bases

The most fundamental concept in transcendental field theory is that of a transcendence basis. In this section, we investigate this concept. We shall see that the notion of a transcendence basis is very similar to that of a basis of a vector space. To give a rough description of a transcendence basis, let K/F be a field extension. A subset T of K is a transcendence basis for K/F if T is a maximal set of "variables" in K. To be a little less vague, $F(T)$ is isomorphic to a rational function field $F(X)$ with $|T| = |X|$, and the maximality means that there is no larger set of variables in K. We need to make this precise, to prove that transcendence bases exist, and to determine their properties.

Definition 19.1 *Let K be a field extension of F, and let $t_1, \ldots, t_n \in K$. The set $\{t_1, \ldots, t_n\}$ is algebraically independent over F if $f(t_1, \ldots, t_n) \neq 0$ for all nonzero polynomials $f \in F[x_1, \ldots, x_n]$. An arbitrary set $S \subseteq K$ is algebraically independent over F if any finite subset of S is algebraically*

independent over F. If a set is not algebraically independent over F, then it is said to be algebraically dependent over F.

Example 19.2 If $K = F(x_1, \ldots, x_n)$ is the field of rational functions over F in n variables, then $\{x_1, \ldots, x_n\}$ is algebraically independent over F. Moreover, if $r_1, \ldots, r_n$ are any positive integers, then $\{x_1^{r_1}, \ldots, x_n^{r_n}\}$ is also algebraically independent over F.

Keeping with the same field extension, let $A = (a_{ij})$ be an $n \times n$ matrix with coefficients in F, and let $f_j = \sum_i a_{ij} x_i$. We prove that $\{f_1, \ldots, f_n\}$ is algebraically independent over F if and only if $\det A \neq 0$. For simplicity, we write $F[X]$ for $F[x_1, \ldots, x_n]$. The matrix A induces a ring homomorphism $A' : F[X] \to F[X]$ that sends x_i to f_i. If $\det A \neq 0$, then A has an inverse; say $A^{-1} = (b_{ij})$, and A^{-1} induces the inverse map $(A^{-1})' : F[X] \to F[X]$ to A'. Therefore, A' is injective, so $h(f_1, \ldots, f_n) \neq 0$ for all nonzero h. Thus, $\{f_1, \ldots, f_n\}$ is algebraically independent over F. Conversely, suppose that $\det A = 0$. Then the columns C_j of A are linearly independent over F; say $\sum_j b_j C_j = 0$ with each $b_j \in F$, and not all of the b_j are zero. A short calculation shows that $\sum_j b_j f_j = 0$; hence, the f_j are algebraically dependent over F.

Example 19.3 By convention, the empty set $\varnothing$ is algebraically independent over any field. The singleton sets $\{e\}$, $\{\pi\}$, and $\{4e^{-1}\}$ are all algebraically independent over $\mathbb{Q}$. The set $\{e, e^2\}$ is not algebraically independent over $\mathbb{Q}$, since $f(e, e^2) = 0$ if $f(x_1, x_2) = x_1^2 - x_2$. It is unknown whether $\{e, \pi\}$ is algebraically independent over $\mathbb{Q}$.

Example 19.4 Let $F \subseteq K \subseteq L$ be fields, and let S be a subset of L. If S is algebraically independent over F, then S is also algebraically independent over K. This is clear from the definition of algebraic independence. Moreover, if T is any subset of S and if S is algebraically independent over F, then T is also algebraically independent over F. The converse of the first statement is false in general. Suppose that $K = F(x) = L$. Then $\{x\}$ is algebraically independent over F, but $\{x\}$ is algebraically dependent over K.

An algebraically independent set of elements behaves the same as a set of variables in a polynomial ring. The following lemma makes this statement precise.

Lemma 19.5 *Let K be a field extension of F. If $t_1, \ldots, t_n \in K$ are algebraically independent over F, then $F[t_1, \ldots, t_n]$ and $F[x_1, \ldots, x_n]$ are F-isomorphic rings, and so $F(t_1, \ldots, t_n)$ and $F(x_1, \ldots, x_n)$ are F-isomorphic fields.*

Proof. Define $\varphi : F[x_1, \ldots, x_n] \longrightarrow K$ by $\varphi(f(x_1, \ldots, x_n)) = f(t_1, \ldots, t_n)$. Then φ is an F-homomorphism of rings. The algebraic independence of

the t_i shows that φ is injective, and the image of φ is $F[t_1, \ldots, t_n]$. Therefore, $F[t_1 \ldots, t_n]$ and $F[x_1, \ldots, x_n]$ are isomorphic. This map induces an F-isomorphism of quotient fields, which finishes the proof. □

Definition 19.6 *A field K is purely transcendental over a subfield F if K is isomorphic to a field of rational functions over F in some number, finite or infinite, of variables.*

If $K = F(t_1, \ldots, t_n)$ with $\{t_1, \ldots, t_n\}$ algebraically independent, then K is often said to be a *rational extension* of F. This terminology is often used in algebraic geometry. We will investigate the geometric significance of rational extensions in Section 22.

We now begin to analyze the definition of algebraic independence.

Lemma 19.7 *Let K be a field extension of F, and let $t_1, \ldots, t_n \in K$. Then the following statements are equivalent:*

1. *The set $\{t_1, \ldots, t_n\}$ is algebraically independent over F.*
2. *For each i, t_i is transcendental over $F(t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n)$.*
3. *For each i, t_i is transcendental over $F(t_1, \ldots, t_{i-1})$.*

Proof. (1) $\Rightarrow$ (2): Suppose that there are $a_j \in F(t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n)$ such that $a_0 + a_1 t_i + \cdots + t_i^m = 0$. We may write $a_j = b_j / c$ with $b_1, \ldots, b_n, c \in F[t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n]$, and so $b_0 + b_1 t_i + \cdots + b_m t_i^m = 0$. If $b_j = g_j(t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n)$, then $f = \sum_j g_j(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) x_i^j$ is a polynomial and $f(t_1, \ldots, t_n) = 0$. Since $\{t_1, \ldots, t_n\}$ is algebraically independent over F, the polynomial f must be 0. Consequently, each $a_j = 0$, so t_i is transcendental over $F(t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n)$.

(2) $\Rightarrow$ (3): If t_i is transcendental over $F(t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n)$, then t_i clearly is transcendental over the smaller field $F(t_1, \ldots, t_{i-1})$.

(3) $\Rightarrow$ (1): Suppose that the t_i are not algebraically independent over F. Choose m minimal such that there is a nonzero $f(x_1, \ldots, x_m) \in F[x_1, \ldots, x_m]$ with $f(t_1, \ldots, t_m) = 0$. Say $f = \sum_j g_j x_m^j$ with $g_j \in F[x_1, \ldots, x_{m-1}]$, and let $a_j = g(t_1, \ldots, t_{m-1})$. Then $a_0 + a_1 t_m + \cdots + a_r t_m^r = 0$. If the a_j are not all zero, then t_m is algebraic over $F(t_1, \ldots, t_m)$, a contradiction. Thus, $a_j = 0$ for each j. By the minimality of m, the $t_1, \ldots, t_{m-1}$ are algebraically independent over F, which implies that all $g_j = 0$, so $f = 0$. This proves that $\{t_1, \ldots, t_n\}$ is algebraically independent over F. □

Definition 19.8 *If K is a field extension of F, a subset S of K is a transcendence basis for K/F if S is algebraically independent over F and if K is algebraic over $F(S)$.*

Example 19.9 If K/F is a field extension, then $\varnothing$ is a transcendence basis for K/F if and only if K/F is algebraic.

Example 19.10 If $K = F(x_1, \ldots, x_n)$, then $\{x_1, \ldots, x_n\}$ is a transcendence basis for K/F. Moreover, if $r_1, \ldots, r_n$ are positive integers, then we show that $\{x_1^{r_1}, \ldots, x_n^{r_n}\}$ is also a transcendence basis for K/F. We saw in Example 19.2 that $\{x_1^{r_1}, \ldots, x_n^{r_n}\}$ is algebraically independent over F. We need to show that K is algebraic over $L = F(x_1^{r_1}, \ldots, x_n^{r_n})$. This is true because for each i the element x_i satisfies the polynomial $t^{r_i} - x_i^{r_i} \in L[t]$.

Here is a natural question that one may have about the definition of transcendence basis: Why is the condition "K is algebraic over $F(S)$" used instead of "$K = F(S)$"? We give two reasons. The previous example shows that even when $K = F(X)$ for some algebraically independent set X over F, there may be other algebraically independent sets Y for which K is algebraic over $F(Y)$ but that $K \neq F(Y)$. Moreover, it is a very restrictive condition to require that a field be purely transcendental over a subfield. Without the definition as it is given, existence of a transcendence basis would be uncommon, and the concept would not be very useful.

The next two examples deal with field extensions of the sort that arise in algebraic geometry. We will study extensions of this type in Section 22.

Example 19.11 Let k be a field, and let $f(x,y) = y^2 - x^3 + x \in k[x,y]$. Then f is an irreducible polynomial, so $A = k[x,y]/(f)$ is an integral domain. Note that A contains an isomorphic copy of k. Let K be the quotient field of $k[x,y]/(f)$. We can then view K as a field extension of k. If $u = x + (f)$ and $v = y + (f)$ are the images of x, y in K, then $K = k(u,v)$. We show that $\{u\}$ is a transcendence basis for K/k. Since $v^2 = u^3 - u$, the field K is algebraic over $k(u)$. We then need to show that u is transcendental over k. If this is false, then u is algebraic over k, so K is algebraic over k. We claim that this forces $A = k[u,v]$ to be a field. To prove this, take $t \in A$. Then $t^{-1} \in K$ is algebraic over k, so $t^{-n} + \alpha_{n-1}t^{n-1} + \cdots + \alpha_0 = 0$ for some $\alpha_i \in k$ with $\alpha_0 \neq 0$. Multiplying by t^{n-1} gives

$$t^{-1} = -(\alpha_{n-1} + \alpha_{n-2}t + \cdots + \alpha_0 t^{n-1}) \in A,$$

proving that A is a field. However, $A = k[x,y]/(f)$ is a field if and only if (f) is a maximal ideal of $k[x,y]$. The ring A cannot be a field, since (f) is properly contained in the ideal (x,y) of $k[x,y]$. Thus, u is not algebraic over k, so $\{u\}$ is a transcendence basis for K/k. Note that a similar argument would show that $\{v\}$ is also a transcendence basis for K/k.

Example 19.12 We give a generalization of the previous example. Let k be a field and let $f \in k[x_1, \ldots, x_n]$ be an irreducible polynomial. Then $A = k[x_1, \ldots, x_n]/(f)$ is an integral domain. Let K be the quotient field of

A. We may write

$$f = g_m x_n^m + g_{m-1} x_n^{m-1} + \cdots + g_0$$

with each $g_i \in k[x_1, \ldots, x_{n-1}]$. Let us assume that $m > 0$, so that f does involve the variable x_n. If $t_i = x_i + (f)$ is the image of x_i in A, we claim that $\{t_1, \ldots, t_{n-1}\}$ is a transcendence basis for K/k. To see this, the equation for f above shows that t_n is algebraic over $k(t_1, \ldots, t_{n-1})$, so we only need to show that $\{t_1, \ldots, t_{n-1}\}$ is algebraically independent over k. Suppose that there is a polynomial $h \in k[x_1, \ldots, x_{n-1}]$ with $h(t_1, \ldots, t_{n-1}) = 0$. Then $h(x_1, \ldots, x_{n-1}) \in (f)$, so f divides h. Thus, $h = fg$ for some $g \in k[x_1, \ldots, x_n]$. However, the polynomial h does not involve the variable x_n while f does, so comparing degrees in x_n of h and fg shows that $h = 0$. Therefore, $\{t_1, \ldots, t_{n-1}\}$ is algebraically independent over k, so $\{t_1, \ldots, t_{n-1}\}$ is a transcendence basis for K/k.

The argument we gave for why $\{t_1, \ldots, t_{n-1}\}$ is algebraically independent over k is different from the argument used in the previous example to show u is transcendental over k. We could have used the argument of this example in the previous example, but we chose to give a different argument to illustrate different methods that can be used in dealing with transcendental extensions.

There is a strong connection between the concepts of linear independence in vector spaces and algebraic independence in fields. In particular, we will prove below that every field extension has a transcendental basis and that the size of a transcendence basis is uniquely determined. The reader would benefit by recalling how the corresponding facts are proved for vector spaces.

Lemma 19.13 *Let K be a field extension of F, and let $S \subseteq K$ be algebraically independent over F. If $t \in K$ is transcendental over $F(S)$, then $S \cup \{t\}$ is algebraically independent over F.*

Proof. Suppose that the lemma is false. Then there is a nonzero polynomial $f \in F[x_1, \ldots, x_n, y]$ with $f(s_1, \ldots, s_n, t) = 0$ for some $s_i \in S$. This polynomial must involve y, since S is algebraically independent over F. Write $f = \sum_{j=0}^{m} g_j y^j$ with $g_j \in F[x_1, \ldots, x_n]$. Since $g_m \neq 0$, the element t is algebraic over $F(S)$, a contradiction. Thus, $S \cup \{t\}$ is algebraically independent over F. □

We now prove the existence of a transcendence basis for any field extension.

Theorem 19.14 *Let K be a field extension of F.*

1. *There exists a transcendence basis for K/F.*

2. *If $T \subseteq K$ such that $K/F(T)$ is algebraic, then T contains a transcendence basis for K/F.*

3. *If $S \subseteq K$ is algebraically independent over F, then S is contained in a transcendence basis of K/F.*

4. *If $S \subseteq T \subseteq K$ such that S is algebraically independent over F and $K/F(T)$ is algebraic, then there is a transcendence basis X for K/F with $S \subseteq X \subseteq T$.*

Proof. We first mention why statement 4 implies the first three statements. If statement 4 is true, then statements 2 and 3 are true by setting $S = \varnothing$ and $T = K$, respectively. Statement 1 follows from statement 4 by setting $S = \varnothing$ and $T = K$. To prove statement 4, let $\mathcal{S}$ be the set of all algebraically independent subsets of T that contain S. Then $\mathcal{S}$ is nonempty, since $S \in \mathcal{S}$. Ordering $\mathcal{S}$ by inclusion, a Zorn's lemma argument shows that $\mathcal{S}$ contains a maximal element M. If K is not algebraic over $F(M)$, then $F(T)$ is not algebraic over $F(M)$, since K is algebraic over $F(T)$. Thus, there is a $t \in T$ with t transcendental over $F(M)$. But by Lemma 19.13, $M \cup \{t\}$ is algebraically independent over F and is a subset of T, contradicting maximality of M. Thus, K is algebraic over $F(M)$, so M is a transcendence basis of K/F contained in X. □

We now show that any two transcendence bases have the same size.

Theorem 19.15 *Let K be a field extension of F. If S and T are transcendence bases for K/F, then $|S| = |T|$.*

Proof. We first prove this in the case where $S = \{s_1, \dots, s_n\}$ is finite. Since S is a transcendence basis for K/F, the field K is not algebraic over $F(S - \{s_1\})$. As K is algebraic over $F(T)$, some $t \in T$ must be transcendental over $F(S - \{s_1\})$. Hence, by Lemma 19.13, $\{s_2, \dots, s_n, t\}$ is algebraically independent over F. Furthermore, s_1 is algebraic over $F(s_2, \dots, s_n, t)$, or else $\{s_1, \dots, s_n, t\}$ is algebraically independent, which is false. Thus, $\{s_2, \dots, s_n, t\}$ is a transcendence basis for K/F. Set $t_1 = t$. Assuming we have found $t_i \in T$ for all i with $1 \le i < m \le n$ such that $\{t_1, \dots, t_{m-1}, s_m, \dots, s_n\}$ is a transcendence basis for K/F, by replacing S by this set, the argument above shows that there is a $t' \in T$ such that $\{t_1, \dots, t_{m-1}, t', \dots, s_n\}$ is a transcendence basis for K/F. Setting $t_m = t'$ and continuing in this way, we get a transcendence basis $\{t_1, \dots, t_n\} \subseteq T$ of K/F. Since T is a transcendence basis for K/F, we see that $\{t_1, \dots, t_n\} = T$, so $|T| = n$.

For the general case, by the previous argument we may suppose that S and T are both infinite. Each $t \in T$ is algebraic over $F(S)$; hence, there is a finite subset $S_t \subseteq S$ with t algebraic over $F(S_t)$. If $S' = \bigcup_{t \in T} S_t$, then each $t \in T$ is algebraic over $F(S')$. Since K is algebraic over $F(T)$, we see

that K is algebraic over $F(S')$. Thus, $S = S'$ since $S' \subseteq S$ and S is a transcendence basis for K/F. We then have

$$|S| = |S'| = |\bigcup_{t \in T} S_t| \leq |T \times \mathbb{N}| = |T|,$$

where the last equality is true since T is infinite. Reversing the argument, we see that $|T| \leq |S|$, so $|S| = |T|$. □

This theorem shows that the size of a transcendence basis for K/F is unique. The following definition is then well defined.

Definition 19.16 *The transcendence degree* $\mathrm{trdeg}(K/F)$ *of a field extension K/F is the cardinality of any transcendence basis of K/F.*

Corollary 19.17 *Let $t_1, \ldots, t_n \in K$. Then the fields $F(t_1, \ldots, t_n)$ and $F(x_1, \ldots, x_n)$ are F-isomorphic if and only if $\{t_1, \ldots, t_n\}$ is an algebraically independent set over F.*

Proof. If $\{t_1, \ldots, t_n\}$ is algebraically independent over F, then $F(t_1, \ldots, t_n)$ and $F(x_1, \ldots, x_n)$ are F-isomorphic fields by Lemma 19.5. Conversely, if $F(t_1, \ldots, t_n) \cong F(x_1, \ldots, x_n)$, suppose that $\{t_1, \ldots, t_n\}$ is algebraically dependent over F. By the previous theorem, there is a subset S of $\{t_1, \ldots, t_n\}$ such that S is a transcendence basis for $F(t_1, \ldots, t_n)/F$. However, the transcendence degree of this extension is n, which forces $|S| = n$, so $S = \{t_1, \ldots, t_n\}$. Thus, $\{t_1, \ldots, t_n\}$ is algebraically independent over F. □

We now prove the main arithmetic fact about transcendence degrees, the following transitivity result.

Proposition 19.18 *Let $F \subseteq L \subseteq K$ be fields. Then*

$$\mathrm{trdeg}(K/F) = \mathrm{trdeg}(K/L) + \mathrm{trdeg}(L/F).$$

Proof. Let S be a transcendence basis for L/F, and let T be a transcendence basis for K/L. We show that $S \cup T$ is a transcendence basis for K/F, which will prove the result because $S \cap T = \varnothing$. Since T is algebraically independent over L, the set T is also algebraically independent over $F(S) \subseteq L$, so $S \cup T$ is algebraically independent over F. To show that K is algebraic over $F(S \cup T)$, we know that $K/L(T)$ and $L/F(S)$ are algebraic. Therefore, $L(T)$ is algebraic over $F(S \cup T) = F(S)(T)$, since each $t \in T$ is algebraic over $F(S \cup T)$. Thus, by transitivity, K is algebraic over $F(S \cup T)$, so $S \cup T$ is a transcendence basis for K/F. This proves the proposition. □

Example 19.19 Let $K = k(x_1, \ldots, x_n)$ be the field of rational functions in n variables over a field k, and let $F = k(s_1, \ldots, s_n)$ be the subfield of K generated over k by the elementary symmetric functions $s_1, \ldots, s_n$. In Example 3.9, we saw that K is an algebraic extension of F with $[K : F] = n!$. Therefore, $\{s_1, \ldots, s_n\}$ contains a transcendence basis of K/k. However, $\{x_1, \ldots, x_n\}$ is a transcendence basis for K/k, so $\operatorname{trdeg}(K/k) = n$. This forces the s_i to be algebraically independent over k; hence, they form a transcendence basis for K/k. In particular, this shows that $k(s_1, \ldots, s_n) \cong k(x_1, \ldots, x_n)$.

Example 19.20 Consider the field extension $\mathbb{C}/\mathbb{Q}$. Since $\mathbb{Q}$ is countable and $\mathbb{C}$ is uncountable, the transcendence degree of $\mathbb{C}/\mathbb{Q}$ must be infinite (in fact, uncountable), for if $t_1, \ldots, t_n$ form a transcendence basis for $\mathbb{C}/\mathbb{Q}$, then $\mathbb{C}$ is algebraic over $\mathbb{Q}(t_1, \ldots, t_n)$, so $\mathbb{C}$ and $\mathbb{Q}$ have the same cardinality, since they are infinite fields. However, one can show that $\mathbb{Q}(t_1, \ldots, t_n)$ is countable. This would give a contradiction to the uncountability of $\mathbb{C}$. Thus, any transcendence basis T of $\mathbb{C}/\mathbb{Q}$ is infinite.

Let T be any transcendence basis of $\mathbb{C}/\mathbb{Q}$. Since $\mathbb{C}$ is algebraic over $\mathbb{Q}(T)$ and is algebraically closed, $\mathbb{C}$ is an algebraic closure of $\mathbb{Q}(T)$. Let σ be a permutation of T. Then σ induces an automorphism of $\mathbb{Q}(T)$ that is trivial on $\mathbb{Q}$; hence, σ extends to an automorphism of $\mathbb{C}$ by the isomorphism extension theorem. Since there are infinitely many such σ, we see that $|\operatorname{Aut}(\mathbb{C})| = \infty$. Because any automorphism of $\mathbb{R}$ is the identity, the only automorphisms of $\mathbb{C}$ that map $\mathbb{R}$ to $\mathbb{R}$ are the identity map and complex conjugation. Thus, there are infinitely many $\sigma \in \operatorname{Aut}(\mathbb{C})$ with $\sigma(\mathbb{R}) \neq \mathbb{R}$. We can easily show that $[\mathbb{C} : \sigma(\mathbb{R})] = 2$. This means that there are infinitely many subfields F of $\mathbb{C}$ with $[\mathbb{C} : F] = 2$. It is a whole different question to try to construct such fields. Note that in order to get these automorphisms of $\mathbb{C}$, we invoked Zorn's lemma twice, once for the existence of a transcendence basis of $\mathbb{C}/\mathbb{Q}$ and the second time indirectly by using the isomorphism extension theorem.

Problems

1. Let K be a field extension of F, let $\alpha \in K$ be algebraic over F, and let $t \in T$ be transcendental over F. Show that $\min(F, \alpha) = \min(F(t), \alpha)$ and that $[F(\alpha) : F] = [F(t, \alpha) : F(t)]$.

2. Suppose that L_1, L_2 are intermediate fields of K/F. Show that
$$\operatorname{trdeg}(L_1L_2/F) \leq \operatorname{trdeg}(L_1/F) + \operatorname{trdeg}(L_2/F).$$

3. Give an example of a field extension K/F with intermediate subfields L_1, L_2 satisfying $\operatorname{trdeg}(L_1L_2/F) < \operatorname{trdeg}(L_1/F) + \operatorname{trdeg}(L_2/F)$.

4. Let K be a finitely generated field extension of F. If L is a field with $F \subseteq L \subseteq K$, show that L/F is finitely generated.

5. Let K be an algebraically closed field, and let F be a subfield of K. If $\varphi : K \to K$ is an F-homomorphism and $\text{trdeg}(K/F) < \infty$, show that φ is surjective, so that φ is an F-automorphism of K.

6. Let K be an algebraically closed field, and let F be a subfield of K with $\text{trdeg}(K/F) = \infty$. Show that there is an F-homomorphism $\varphi : K \to K$ that is not an F-automorphism.

7. Let $K = \mathbb{C}(x)(\sqrt{-1-x^2})$. Show that $[K : \mathbb{C}(x)] = 2$, and show that $K = \mathbb{C}(t)$ if $t = (i-x)^{-1}\sqrt{(-1-x^2)}/(i-x)$.

8. Let $K = \mathbb{R}(x)(\sqrt{-1-x^2})$. Show that $[K : \mathbb{R}(x)] = 2$ and that there is no $t \in K$ with $K = \mathbb{R}(t)$.

9. If $K = \mathbb{R}(x)(\sqrt{1+x^2})$, show that there is a $t \in K$ with $K = \mathbb{R}(t)$.

10. Let x be transcendental over $\mathbb{C}$, and let K be the algebraic closure of $\mathbb{C}(x)$. Prove that $K \cong \mathbb{C}$.

11. Let $K = F(x)$ be the rational function field over a field F of characteristic 0, let $L_1 = F(x^2)$, and let $L_2 = F(x^2 + x)$.

    (a) Show that $|\text{Gal}(K/L_i)| = 2$ for each i, and find the unique non-identity L_i-automorphism of K.

    (b) Show that $L_1 \cap L_2 = F$.
    (Hint: What is the subgroup of $\text{Gal}(K/F)$ generated by the automorphisms in the first part?)

12. Let $F(x)$ be the rational function field in one variable over a field F. Show that

$$[F(x) : F] = \begin{cases} |F| & \text{if } F \text{ is infinite,} \\ \aleph_0 & \text{if } F \text{ is finite.} \end{cases}$$

In the following problems, we axiomatize the properties common to linear dependence and algebraic independence, and we see that these two situations can be analyzed simultaneously.

Let X be a set, and let $\prec$ be a relation between elements of X and subsets of X. We will write $\alpha \prec S$ if the relation holds between $\alpha \in X$ and $S \subseteq X$. The relation $\prec$ is called a *dependence relation* if the following conditions hold: (i) if $\alpha \in S$, then $\alpha \prec S$; (ii) if $\alpha \prec S$, then there is a finite subset S_0 of S with $\alpha \prec S$; (iii) if T is a set such that $s \prec T$ for all $s \in S$, and if $\alpha \prec S$, then $\alpha \prec T$; and (iv) if $\alpha \prec S$ but $\alpha \not\prec S - \{s\}$ for some $s \in S$, then $s \prec (S - \{s\}) \cup \{\alpha\}$.

If $\prec$ is a dependence relation on X, a subset S of X is *independent* if $s \not\prec S - \{s\}$ for all $s \in S$. If $S \subseteq T$, we say that S *spans* T if $t \prec S$ for each

$t \in T$. Finally, we say that S is a *basis* of X if S is both independent and spans X.

13. Let F be a field, and let V be an F-vector space. Define $\prec$ by $v \prec S$ if v is in the subspace spanned by S. Show that $\prec$ is a dependence relation on V.

14. Let K be a field extension of F. Define $\prec$ by $a \prec S$ if a is algebraic over $F(S)$. Show that $\prec$ is a dependence relation on K.

15. Let K be a field extension of F. Define $\prec$ by $a \prec S$ if $a \in F(S)$. Is $\prec$ a dependence relation on K?

16. Let K/F be a field extension with $\text{char}(F) = p > 0$. Suppose that $K^p \subseteq F$. For instance, we could take K to be any field of characteristic p and $F = K^p$. Define $\prec$ by $a \prec S$ if $a \in F(S)$. Show that $\prec$ is a dependence relation. This is called the relation of *p-dependence*. This relation will show up in Section 23.

17. In this problem, we outline a proof that a set X with a dependence relation $\prec$ has a basis. Prove the following statements.

    (a) Let $S \subseteq T$ be subsets of X, and let $\alpha \in X$. If $\alpha \prec S$, show that $\alpha \prec T$. Conclude that if S is independent, then any subset of S is independent, and if T spans X, then any set containing T also spans X.

    (b) If S is independent and $\alpha \not\prec S$, show that $S \cup \{\alpha\}$ is independent.

    (c) If $S \subseteq T$ are subsets of X such that S is independent and T spans X, show that there is a basis B of X with $S \subseteq B \subseteq T$.

18. In this problem, we show that any two bases of a set X with a dependence relation $\prec$ have the same size. Mimic the proofs of the appropriate results of the section to verify the following steps.

    (a) Suppose that B is a basis of X. If $\beta \in B$ and $\alpha \in X$, let $B' = (B - \{\beta\}) \cup \{\alpha\}$. If $B \prec B'$, then show that B' is also a basis of X.

    (b) If B and C are bases of X with $|B|$ finite, show that $|B| = |C|$.

    (c) If B and C are bases of X, show that $|B| = |C|$.

20 Linear Disjointness

In this section, we study linear disjointness, a technical condition but one with many applications. One way that we use this concept is to extend

the definition of separability in a useful way to nonalgebraic extensions. We tacitly assume that all of our field extensions of a given field F lie in some common extension field C of F. Problem 6 shows that this is not a crucial assumption. We will also make use of tensor products. By phrasing some results in terms of tensor products, we are able to give cleaner, shorter proofs. However, the basic results on linear disjointness can be proved without using tensor products. Properties of tensor products are given in Appendix D for the benefit of the reader.

Definition 20.1 *Let K and L be subfields of a field C, each containing a field F. Then K and L are linearly disjoint over F if every F-linearly independent subset of K is also linearly independent over L.*

Let A and B be subrings of a commutative ring R. Then the ring $A[B]$ is the subring of R generated by A and B; that is, $A[B]$ is the smallest subring of R containing $A \cup B$. It is not hard to show that

$$A[B] = \left\{ \sum a_i b_i : a_i \in A, b_i \in B \right\}.$$

If A and B contain a common field F, then the universal mapping property of tensor products shows that there is a well-defined F-linear transformation $\varphi : A \otimes_F B \to A[B]$ given on generators by $\varphi(a \otimes b) = ab$. We refer to the map φ as the natural map from $A \otimes_F B$ to $A[B]$. We give a criterion in terms of tensor products for two fields to be linear disjoint over a common subfield.

Proposition 20.2 *Let K and L be field extensions of a field F. Then K and L are linearly disjoint over F if and only if the map $\varphi : K \otimes_F L \to K[L]$ given on generators by $a \otimes b \mapsto ab$ is an isomorphism of F-vector spaces.*

Proof. The natural map $\varphi : K \otimes_F L \to K[L]$ is surjective by the description of $K[L]$ given above. So, we need to show that K and L are linearly disjoint over F if and only if φ is injective. Suppose first that K and L are linearly disjoint over F. Let $\{k_i\}_{i \in I}$ be a basis for K as an F-vector space. Each element of $K \otimes_F L$ has a unique representation in the form $\sum k_i \otimes l_i$, with the $l_i \in L$. Suppose that $\sum k_i \otimes l_i \in \ker(\varphi)$, so $\sum k_i l_i = 0$. Then each $l_i = 0$, since K and L are linearly disjoint over F and $\{k_i\}$ is F-linearly independent. Thus, φ is injective, and so φ is an isomorphism.

Conversely, suppose that the map φ is an isomorphism. Let $\{a_j\}_{j \in J}$ be an F-linearly independent subset of K. By enlarging J, we may assume that the set $\{a_j\}$ is a basis for K. If $\{a_j\}$ is not L-linearly independent, then there are $l_j \in L$ with $\sum a_j l_j = 0$, a finite sum. Then $\sum a_j \otimes l_j \in \ker(\varphi)$, so $\sum a_j \otimes l_j = 0$ by the injectivity of φ. However, elements of $K \otimes_F L$ can be represented uniquely in the form $\sum a_j \otimes m_j$ with $m_j \in L$. Therefore, each $l_j = 0$, which forces the set $\{a_j\}$ to be L-linearly independent. Thus, K and L are linearly disjoint over F. □

Corollary 20.3 *The definition of linear disjointness is symmetric; that is, K and L are linearly disjoint over F if and only if L and K are linearly disjoint over F.*

Proof. This follows from Proposition 20.2. The map $\varphi : K \otimes_F L \to K[L]$ is an isomorphism if and only if $\tau : L \otimes_F K \to L[K] = K[L]$ is an isomorphism, since $\tau = i \circ \varphi$, where i is the canonical isomorphism $K \otimes_F L \to L \otimes_F K$ that sends $a \otimes b$ to $b \otimes a$. □

Lemma 20.4 *Suppose that K and L are finite extensions of F. Then K and L are linearly disjoint over F if and only if $[KL : F] = [K : F] \cdot [L : F]$.*

Proof. The natural map $\varphi : K \otimes_F L \to K[L]$ that sends $k \otimes l$ to kl is surjective and

$$\dim(K \otimes_F L) = [K : F] \cdot [L : F].$$

Thus, φ is an isomorphism if and only if $[KL : F] = [K : F] \cdot [L : F]$. The lemma then follows from Proposition 20.2. □

Example 20.5 Suppose that K and L are extensions of F with $[K : F]$ and $[L : F]$ relatively prime. Then K and L are linearly disjoint over F. To see this, note that both $[K : F]$ and $[L : F]$ divide $[KL : F]$, so their product divides $[KL : F]$ since these degrees are relatively prime. The linear disjointness of K and L over F follows from the lemma.

Example 20.6 Let K be a finite Galois extension of F. If L is any extension of F, then K and L are linearly disjoint over F if and only if $K \cap L = F$. This follows from the previous example and the theorem of natural irrationalities, since

$$[KL : F] = [L : F][K : K \cap L],$$

so $[KL : F] = [K : F][L : F]$ if and only if $K \cap L = F$.

The tensor product characterization of linear disjointness leads us to believe that there is a reasonable notion of linear disjointness for rings, not just fields. Being able to discuss linear disjointness in the case of integral domains will make it easier to work with fields, as we will see in Section 22 and later in this section.

Definition 20.7 *Let A and B be subrings of a field C, each containing a field F. Then A and B are linearly disjoint over F if the natural map $A \otimes_F B \to C$ given by $a \otimes b \mapsto ab$ is injective.*

Lemma 20.8 *Suppose that F is a field, and $F \subseteq A \subseteq A'$ and $F \subseteq B \subseteq B'$ are all subrings of a field C. If A' and B' are linearly disjoint over F, then A and B are linearly disjoint over F.*

Proof. This follows immediately from properties of tensor products. There is a natural injective homomorphism $i : A \otimes_F B \to A' \otimes_F B'$ sending $a \otimes b$ to $a \otimes b$ for $a \in A$ and $B \in B$. If the natural map $\varphi' : A' \otimes_F B' \to A'[B']$ is injective, then restricting φ to the image of i shows that the map $\varphi : A \otimes_F B \to A[B]$ is also injective. □

Example 20.9 Let K and L be extensions of a field F. If $K \cap L$ is larger than F, then K and L are not linearly disjoint over F by the preceding lemma since $K \cap L$ is not linearly disjoint to itself over F. However, K and L may not be linearly disjoint over F even if $K \cap L = F$. As an example, let $F = \mathbb{Q}$, $K = F(\sqrt[3]{2})$, and $L = F(\omega\sqrt[3]{2})$, where ω is a primitive third root of unity. Then $K \cap L = F$, but $KL = F(\sqrt[3]{2}, \omega)$ has dimension 6 over F, whereas $K \otimes_F L$ has dimension 9, so the map $K \otimes_F L \to KL$ is not injective.

Lemma 20.10 *Suppose that A and B are subrings of a field C, each containing a field F, with quotient fields K and L, respectively. Then A and B are linearly disjoint over F if and only if K and L are linearly disjoint over F.*

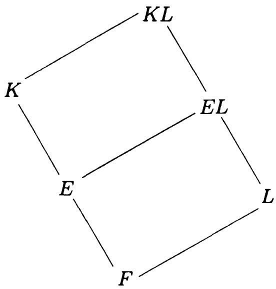
Proof. If K and L are linearly disjoint over F, then A and B are also linearly disjoint over F by the previous lemma. Conversely, suppose that A and B are linearly disjoint over F. Let $\{k_1, \ldots, k_n\} \subseteq K$ be an F-linearly independent set, and suppose that there are $l_i \in L$ with $\sum k_i l_i = 0$. There are nonzero $s \in A$ and $t \in B$ with $sk_i \in A$ and $tl_i \in B$ for each i. The set $\{a_1, \ldots, a_n\}$ is also F-linearly independent; consequently, $\sum a_i \otimes b_i \neq 0$, since it maps to the nonzero element $\sum a_i \otimes b_i \in K \otimes_F L$ under the natural map $A \otimes_F B \to K \otimes_F B$. However, $\sum a_i \otimes b_i$ is in the kernel of the map $A \otimes_F B \to A[B]$; hence, it is zero by the assumption that A and B are linearly disjoint over F. This shows that $\{k_i\}$ is L-linearly independent, so K and L are linearly disjoint over F. □

Example 20.11 Suppose that K/F is an algebraic extension and that L/F is a purely transcendental extension. Then K and L are linearly disjoint over F; to see this, let X be an algebraically independent set over F with $L = F(X)$. From the previous lemma, it suffices to show that K and $F[X]$ are linearly disjoint over F. We can view $F[X]$ as a polynomial ring in the variables $x \in X$. The ring generated by K and $F[X]$ is the polynomial ring $K[X]$. The standard homomorphism $K \otimes_F F[X] \to K[X]$ is an isomorphism because there is a ring homomorphism $\tau : K[X] \to K \otimes_F F[X]$ induced by $x \mapsto 1 \otimes x$ for each $x \in X$, and this is the inverse of φ. Thus, K and $F[X]$ are linearly disjoint over F, so K and L are linearly disjoint over F.

The following theorem is a transitivity property for linear disjointness.

Theorem 20.12 *Let K and L be extension fields of F, and let E be a field with $F \subseteq E \subseteq K$. Then K and L are linearly disjoint over F if and only if E and L are linearly disjoint over F and K and EL are linearly disjoint over E.*

Proof. We have the following tower of fields.



Consider the sequence of homomorphisms

$$K \otimes_F L \xrightarrow{f} K \otimes_E (E \otimes_F L) \xrightarrow{\varphi_1} K \otimes_E EL \xrightarrow{\varphi_2} K[L],$$

where the maps f, φ_1, and φ_2 are given on generators by

$$\begin{aligned} f(k \otimes l) &= k \otimes (1 \otimes l), \\ \varphi_1(k \otimes (e \otimes l)) &= k \otimes el, \\ \varphi_2(k \otimes \sum e_i l_i) &= \sum k e_i l_i, \end{aligned}$$

respectively. Each can be seen to be well defined by the universal mapping property of tensor products. The map f is an isomorphism by counting dimensions. Moreover, φ_1 and φ_2 are surjective. The composition of these three maps is the standard map $\varphi : K \otimes_F L \to K[L]$. First, suppose that K and L are linearly disjoint over F. Then φ is an isomorphism by Proposition 20.2. This forces both φ_1 and φ_2 to be isomorphisms, since all maps in question are surjective. The injectivity of φ_2 implies that K and EL are linearly disjoint over E. If $\sigma : E \otimes_F L \to E[L]$ is the standard map, then φ_1 is given on generators by $\varphi_1(k \otimes (e \otimes l)) = k \otimes \sigma(e \otimes l)$; hence, σ is also injective. This shows that E and L are linearly disjoint over F.

Conversely, suppose that E and L are linearly disjoint over F and that K and EL are linearly disjoint over E. Then φ_2 and σ are isomorphisms by Proposition 20.2. The map φ_1 is also an isomorphism; this follows from the relation between φ_1 and σ above. Then φ is a composition of three isomorphisms; hence, φ is an isomorphism. Using Proposition 20.2 again, we see that K and L are linearly disjoint over F. □

Separability of field extensions

One of the benefits of discussing linear disjointness is that it allows us to give a meaningful notion of separability for arbitrary field extensions. In

Section 22, we shall see some geometric consequences of this more general notion of separability. We first give an example that will help to motivate the definition of separability for nonalgebraic extensions.

Example 20.13 Let K/F be a separable extension, and let L/F be a purely inseparable extension. Then K and L are linearly disjoint over F. To prove this, note that if $\mathrm{char}(F) = 0$, then $L = F$, and the result is trivial. So, suppose that $\mathrm{char}(F) = p > 0$. We first consider the case where K/F is a finite extension. By the primitive element theorem, we may write $K = F(a)$ for some $a \in K$. Let $f(x) = \min(F, a)$ and $g(x) = \min(L, a)$. Then g divides f in $L[x]$. If $g(x) = \alpha_0 + \cdots + \alpha_{n-1}x^{n-1} + x^n$, then for each i there is a positive integer r_i with $\alpha_i^{p^{r_i}} \in F$. If r is the maximum of the r_i, then $\alpha_i^{p^r} \in F$ for each i, so $g(x)^{p^r} \in F[x]$. Consequently, $g(x)^{p^r}$ is a polynomial over F for which a is a root. Thus, f divides g^{p^r} in $F[x]$. Viewing these two divisibilities in $L[x]$, we see that the only irreducible factor of f in $L[x]$ is g, so f is a power of g. The field extension K/F is separable; hence, f has no irreducible factors in any extension field of F. This forces $f = g$, so

$$\begin{aligned} [KL : L] &= [L(a) : L] = \deg(g) \\ &= \deg(f) = [K : F]. \end{aligned}$$

From this, we obtain $[KL : F] = [K : F] \cdot [L : F]$, so K and L are linearly disjoint over F by Lemma 20.4.

If K/F is not necessarily finite, suppose that $\varphi : K \otimes_F L \to KL$ is not injective. Then there are $k_1, \ldots, k_n \in K$ and $l_1, \ldots, l_n \in L$ with $\varphi(\sum k_i \otimes l_i) = 0$. If K_0 is the field generated over F by the k_i, then the restriction of φ to $K_0 \otimes_F L$ is not injective, which is false by the finite dimensional case. Thus, φ is injective, so K and L are linearly disjoint over F.

Definition 20.14 *Let F be a field of characteristic $p > 0$, and let F_{ac} be an algebraic closure of F. Let*

$$F^{1/p^n} = \left\{ a \in F_{ac} : a^{p^n} \in F \right\}$$

and

$$\begin{aligned} F^{1/p^\infty} &= \left\{ a \in F_{ac} : a^{p^n} \in F \textit{ for some } n \geq 0 \right\} \\ &= \bigcup_{n=1}^{\infty} F^{1/p^n}. \end{aligned}$$

The field F^{1/p^∞} is the composite of all purely inseparable extensions of F in F_{ac}. It is, therefore, the maximal purely inseparable extension of F in F_{ac}, so F^{1/p^∞} is the purely inseparable closure of F in F_{ac}.

Definition 20.15 *A transcendence basis X for a field extension K/F is said to be a separating transcendence basis for K/F if K is separable algebraic over $F(X)$. If K has a separating transcendence basis over F, then K is said to be separably generated over F.*

Example 20.16 Let $K = F(x)$ be the rational function field in one variable over a field F of characteristic p. Then $\{x\}$ is a separating transcendence basis for K/F. However, $\{x^p\}$ is also a transcendence basis, but $K/F(x^p)$ is not separable. This example shows that even if K/F is separably generated, not all transcendence bases of K/F are separating transcendence bases.

Example 20.17 If K/F is algebraic, then K is separable over F if and only if K/F is separably generated, so the definition of separably generated agrees with the definition of separable for algebraic extensions.

We now prove the result that characterizes separability of arbitrary extensions.

Theorem 20.18 *Let K be a field extension of F. Then the following statements are equivalent:*

1. *Every finitely generated subextension of K/F is separably generated.*
2. *The fields K and F^{1/p^∞} are linearly disjoint over F.*
3. *The fields K and $F^{1/p}$ are linearly disjoint over F.*

Proof. (1) $\Rightarrow$ (2): To show that K and F^{1/p^∞} are linearly disjoint over F, it suffices to assume that K is a finitely generated extension of F. By statement 1, we know that K is separably generated over F, so there is a transcendence basis $\{t_1, \ldots, t_n\}$ of K/F for which K is separable over $F(t_1, \ldots, t_n)$. By Example 20.11, the fields $F(t_1, \ldots, t_n)$ and F^{1/p^∞} are linearly disjoint over F. Also, K and $F^{1/p^\infty}(t_1, \ldots, t_n)$ are linearly disjoint over $F(t_1, \ldots, t_n)$ by Example 20.13, since $F^{1/p^\infty}(t_1, \ldots, t_n)$ is purely inseparable over $F(t_1, \ldots, t_n)$ and K is separable over $F(t_1, \ldots, t_n)$. Therefore, by Theorem 20.12, the fields K and F^{1/p^∞} are linearly disjoint over F.

(2) $\Rightarrow$ (3): This is clear since $F^{1/p}$ is a subfield of F^{1/p^∞}.

(3) $\Rightarrow$ (1): Suppose that K and $F^{1/p}$ are linearly disjoint over F. Let $L = F(a_1, \ldots, a_n)$ be a finitely generated subextension of K. We use induction on n to show that $\{a_1, \ldots, a_n\}$ contains a separating transcendence basis for L/F. The case $n = 0$ is clear, as is the case where $\{a_1, \ldots, a_n\}$ is algebraically independent, since then $\{a_1, \ldots, a_n\}$ is a separating transcendence basis for L/F. We may then assume that $n > 0$ and that $\{a_1, \ldots, a_m\}$ is a transcendence basis for L/F, with $m < n$. The elements $a_1, \ldots, a_{m+1}$

are algebraically dependent over F, so there is a nonzero polynomial $f \in F[x_1, \dots, x_{m+1}]$ of least total degree with $f(a_1, \dots, a_{m+1}) = 0$. The assumption that f is chosen of least degree forces f to be irreducible. We first claim that f is not a polynomial in $x_1^p, \dots, x_{m+1}^p$. If $f(x_1, \dots, x_{m+1}) = g(x_1^p, \dots, x_{m+1}^p)$ for some $g \in F[x_1, \dots, x_{m+1}]$, then there is an $h \in F^{1/p}[x_1, \dots, x_{m+1}]$ with $f = h(x_1, \dots, x_{m+1})^p$, since we are assuming that $\mathrm{char}(F) = p$ and every coefficient of g is a pth power in $F^{1/p}$. But this implies that $h(a_1, \dots, a_{m+1}) = 0$. Write $h(x_1, \dots, x_{m+1}) = \sum_j \alpha_j m_j$, where the m_j are the monomials occurring in h and the $\alpha_j \in F^{1/p}$. Then $\sum_j \alpha_j m_j(a_1, \dots, a_{m+1}) = 0$, so the $m_j(a_1, \dots, a_{m+1})$ are linearly dependent over $F^{1/p}$. However, since each m_j is a monomial in the x_k, each $m_j(a_1, \dots, a_{m+1}) \in L \subseteq K$. The assumption that K and $F^{1/p}$ are linearly disjoint over F then forces the $m_j(a_1, \dots, a_{m+1})$ to be linearly dependent over F. If $\sum_j \beta_j m_j(a_1, \dots, a_{m+1}) = 0$ with $\beta_j \in F$, then $h' = \sum_j \beta_j m_j$ is a polynomial with $h'(a_1, \dots, a_{m+1}) = 0$ and $\deg(h') < \deg(f)$. This contradiction verifies our claim that f is not a polynomial in $x_1^p, \dots, x_{m+1}^p$. Therefore, for some i the polynomial f is not a polynomial in x_i^p. Let

$$\begin{aligned} q(t) &= f(a_1, \dots, a_{i-1}, t, a_{i+1}, \dots, a_{m+1}) \\ &\in F[a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_{m+1}][t]. \end{aligned}$$

Then $q(a_i) = 0$, and q is not a polynomial in t^p. If we can show that q is irreducible over M, then we will have proved that a_i is separable over M. To see this, the set $\{a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_{m+1}\}$ is a transcendence basis for L/F, so

$$\begin{aligned} F[x_1, \dots, x_{m+1}] &\cong F[a_1, \dots, a_{i-1}, t, a_{i+1}, \dots, a_{m+1}] \\ &= F[a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_{m+1}][t] \end{aligned}$$

as rings. Under the map that sends a_j to x_j and t to x_i, the polynomial q is mapped to f. But f is irreducible over F, so q is irreducible in $F[a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_{m+1}][t]$. By Gauss' lemma, this means that q is irreducible over M, the quotient field of $F[a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_{m+1}]$. Thus, we have shown that a_i is separable over M, so a_i is separable over $L' = F(a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n)$. The induction hypothesis applied to L' gives us a subset of $\{a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n\}$ that is a separating transcendence basis for L'/F. Since a_i is separable over L', this is also a separating transcendence basis for L/F. □

Definition 20.19 *A field extension K/F is separable if* $\mathrm{char}(F) = 0$ *or if* $\mathrm{char}(F) = p > 0$ *and the conditions in Theorem* 20.18 *are satisfied; that is, K/F is separable if every finitely generated subextension of K/F is separably generated.*

We now give some immediate consequences of Theorem 20.18.

Corollary 20.20 *If K/F is separably generated, then K/F is separable. Conversely, if K/F is separable and finitely generated, then K/F is separably generated.*

Corollary 20.21 *Suppose that $K = F(a_1, \ldots, a_n)$ is finitely generated and separable over F. Then there is a subset Y of $\{a_1, \ldots, a_n\}$ that is a separating transcendence basis of K/F.*

Proof. This corollary is more accurately a consequence of the proof of (3) $\Rightarrow$ (1) in Theorem 20.18, since the argument of that step is to show that if K is finitely generated over F, then any finite generating set contains a separating transcendence basis. □

Corollary 20.22 *Let F be a perfect field. Then any finitely generated extension of F is separably generated.*

Proof. This follows immediately from part 3 of Theorem 20.18, since $F^{1/p^\infty} = F$ if F is perfect. □

Corollary 20.23 *Let $F \subseteq E \subseteq K$ be fields.*

1. *If K/F is separable, then E/F is separable.*
2. *If E/F and K/E are separable, then K/F is separable.*
3. *If K/F is separable and E/F is algebraic, then K/E is separable.*

Proof. Part 1 is an immediate consequence of condition 2 of Theorem 20.18. For part 2 we use Theorems 20.18 and 20.12. If E/F and K/E are separable, then E and $F^{1/p}$ are linearly disjoint over F, and K and $E^{1/p}$ are linearly disjoint over E. However, it follows from the definition that $F^{1/p} \subseteq E^{1/p}$, so $EF^{1/p} \subseteq E^{1/p}$. Thus, K and $EF^{1/p}$ are linearly disjoint over E. Theorem 20.12 then shows that K and $F^{1/p}$ are linearly disjoint over F, so K is separable over F.

To prove part 3, suppose that K/F is separable and E/F is algebraic. We know that E/F is separable by part 1. Let $L = E(a_1, \ldots, a_n)$ be a finitely generated subextension of K/E. If $L' = F(a_1, \ldots, a_n)$, then by the separability of K/F there is a separating transcendence basis $\{t_1, \ldots, t_m\}$ for L'/F. Because E/F is separable algebraic, $EL' = L$ is separable over L', so by transitivity, L is separable over $F(t_1, \ldots, t_m)$. Thus, L is separable over $E(t_1, \ldots, t_m)$, so $\{t_1, \ldots, t_m\}$ is a separating transcendence basis for L/E. We have shown that L/E is separably generated for every finitely generated subextension of K/E, which proves that K/E is separable. □

Example 20.24 Let F be a field of characteristic p, let $K = F(x)$, the rational function field in one variable over F, and let $E = F(x^p)$. Then K/F is separable, but K/E is not separable. This example shows the necessity for the assumption that E/F be algebraic in the previous corollary.

Example 20.25 Here is an example of a separable extension that is not separably generated. Let F be a field of characteristic p, let x be transcendental over F, and let $K = F(x)(\{x^{1/p^n} : n \geq 1\})$. Then K is the union of the fields $F(x^{1/p^n})$, each of which is purely transcendental over F, and hence is separably generated. Any finitely generated subextension E is a subfield of $F(x^{1/p^n})$ for some n and hence is separably generated over F by the previous corollary. Therefore, K/F is separable. But K is not separably generated over F, since given any $f \in K$, there is an n with $f \in F(x^{1/p^n})$, so $K/F(f)$ is not separable, since $K/F(x^{1/p^n})$ is a nontrivial purely inseparable extension.

Problems

1. Let F be a field. Show that every field extension of F is separable if and only if F is perfect.

2. Let $\{x, y\}$ be algebraically independent over F. Show that $F(x)$ and $F(y)$ are linearly disjoint over F.

3. Let F be a perfect field, and let K/F be a field extension of transcendence degree 1. If K is not perfect, show that K/F is separably generated.
   (Note: The field K of Example 20.25 is perfect.)

4. Let F be a field, and let F_{ac} be an algebraic closure of F. Then the *perfect closure* of F is the smallest subfield of F_{ac} containing F that is perfect. Show that F^{1/p^∞} is the perfect closure of F.

5. Prove or disprove: Let K be a finite extension of F, and let L be a field extension of F such that K and L are linearly disjoint over F. If N is the normal closure of K/F, then N and L are linearly disjoint over F.

6. Let K and L be two field extensions of F. Show that there is a field extension C of F that contains F-isomorphic copies of both K and L.

7. Let K and L be extensions of a field F. Then K and L are said to be *free* over F if every subset of K that is algebraically independent over F is also algebraically independent over L.

   (a) Show that this definition is symmetric; that is, show that K and L are free over F if and only if L and K are free over F.

   (b) Show that there exists a field extension M of F that contains F-isomorphic copies K' and L' of K and L, respectively, such that (i) M is the composite of K' and L' and (ii) K' and L' are free over F.
   (The field M is called the *free join* of K and L over F.)

8. Let K and L be extensions of a field F. If K and L are linearly disjoint over F, show that K and L are free over F. Give an example to show that the converse is false.

9. Let K be a separable extension of F. If L is an extension of F, show that KL/L is separable, provided that K and L are free over F. Give an example to show that this can be false if K and L are not free over F.

10. Let K/F and L/F be separable extensions. Show that KL/F is separable, provided that K and L are free over F. Give an example to show this can be false if K and L are not free over F.

21 Algebraic Varieties

Field extensions that are finitely generated but not algebraic arise naturally in algebraic geometry. In this section, we discuss some of the basic ideas of algebraic geometry, and in Section 22 we describe the connection between varieties and finitely generated field extensions.

Let k be a field, and let $f \in k[x_1, \dots, x_n]$ be a polynomial in the n variables $x_1, \dots, x_n$. Then f can be viewed as a function from k^n to k in the obvious way; if $P = (a_1, \dots, a_n) \in k^n$, we will write $f(P)$ for $f(a_1, \dots, a_n)$. It is possible for two different polynomials to yield the same function on k^n. For instance, if $k = \mathbb{F}_2$, then $x^2 - x$ is the zero function on k^1, although it is not the zero polynomial. However, if k is infinite, then $f \in k[x_1, \dots, x_n]$ is the zero function on k^n if and only if f is the zero polynomial (see Problem 1).

Definition 21.1 *Let k be a field, and let C be an algebraically closed field containing k. If S is a subset of $k[x_1, \dots, x_n]$, then the zero set of S is*

$$Z(S) = \{(a_1, \dots, a_n) \in C^n : f(a_1, \dots, a_n) = 0 \text{ for all } f \in S\}.$$

Definition 21.2 *Let k be a field, and let C be an algebraically closed field containing k. Then a set $V \subseteq C^n$ is said to be a k-variety if $V = Z(S)$ for some set S of polynomials in $k[x_1, \dots, x_n]$. The set*

$$V(k) = \{P \in k^n : f(P) = 0 \text{ for all } f \in S\}$$

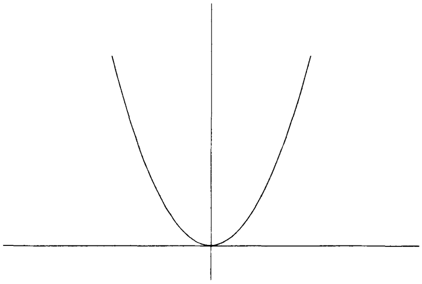
is called the set of k-rational points of V.

Before looking at a number of examples, we look more closely at the definitions above. The reason for working in C^n instead of k^n is that a polynomial $f \in k[x_1, \dots, x_n]$ may not have a zero in k^n, but, as we shall see below, f does have zeros in C^n. For example, if $f = x^2 + y^2 + 1 \in \mathbb{R}[x, y]$,

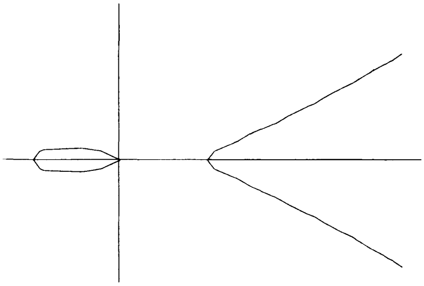
then f has no zeros in $\mathbb{R}^2$, while f has the zeros $(0, \pm i)$, among others, in $\mathbb{C}^2$. Classical algebraic geometry is concerned with polynomials over $\mathbb{C}$. On the other hand, zeros of polynomials over a number field are of concern in algebraic number theory. Working with polynomials over a field k but looking at zeros inside C^n allows one to handle both of these situations simultaneously.

We now look at some examples of varieties. The pictures below show the $\mathbb{R}$-rational points of the given varieties.

Example 21.3 Let $f(x,y) = y - x^2$. Then $Z(f) = \{(a, a^2) : a \in C\}$, a k-variety for any $k \subseteq C$.



Example 21.4 Let $f(x,y) = y^2 - (x^3 - x)$. Then $Z(f)$ is a k-variety for any $k \subseteq C$. This variety is an example of an *elliptic curve*, a class of curves of great importance in number theory.



Example 21.5 Let $f(x,y) = x^n + y^n - 1 \in \mathbb{Q}[x,y]$, the *Fermat curve.* Fermat's last theorem states that if $V = Z(f)$ and $n \geq 3$, then V has no $\mathbb{Q}$-rational points other than the "trivial points," when either $x = 0$ or $y = 0$.

Example 21.6 Let $V = \{(t^2, t^3) : t \in C\}$. Then V is the k-variety $Z(y^2 - x^3)$. The description of V as the set of points of the form (t^2, t^3) is called a *parameterization* of V. We will see a connection between parameterizing varieties and field extensions in Section 22.

Example 21.7 Let $V = \{(t^3, t^4, t^5) : t \in C\}$. Then V is a k-variety, since V is the zero set of $\{y^2 - xz, z^2 - x^2y\}$. To verify this, note that each point of V does satisfy these two polynomials. Conversely, suppose that $(a, b, c) \in C^3$ is a zero of these three polynomials. If $a = 0$, then a quick check of the polynomials shows that $b = c = 0$, so $(a, b, c) \in V$. If $a \neq 0$, then define $t = b/a$. From $b^2 = ac$, we see that $c = t^2a$. Finally, the equation $c^2 = a^2b$ yields $t^4a^2 = a^3t$, so $a = t^3$. Thus, $(a, b, c) = (t^3, t^4, t^5) \in V$.

Example 21.8 Let $S^n = \{(a_1, \dots, a_n) \in C^n : \sum_{i=1}^n a_i^2 = 1\}$. Then $V = Z(-1 + \sum_{i=1}^n x_i^2)$, so V is a k-variety.

Example 21.9 Let V be a C-vector subspace of C^n. We can find a matrix A such that V is the nullspace of A. If $A = (\alpha_{ij})$, then a point $(a_1, \dots, a_n)$ is in V if and only if $\sum_j \alpha_{ij}a_j = 0$ for each i. Thus, V is the zero set of the set of linear polynomials $\sum_j \alpha_{ij}x_j$, so V is a C-variety. If each α_{ij} lies in a subfield k, then V is a k-variety.

Example 21.10 Let $SL_n(C)$ be the set of all $n \times n$ matrices over C of determinant 1. We view the set of all $n \times n$ matrices over C as the set C^{n^2} of n^2-tuples over C. The determinant $\det = \det(x_{ij})$ is a polynomial in the n^2 variables x_{ij}, and the coefficients of the determinant polynomial are ± 1. We then see that $SL_n(C) = Z(\det - 1)$ is a k-variety for any subfield k of C. For instance, if $n = 2$, then

$$SL_2(C) = \{(a, b, c, d) \in C^4 : ad - bc - 1 = 0\}.$$

We can define a topology on C^n, the *k-Zariski topology*, by defining a subset of C^n to be closed if it is a k-variety. The following lemma shows that this does indeed define a topology on C^n. Some of the problems below go into more detail about the k-Zariski topology.

Lemma 21.11 *The sets $\{Z(S) : S \subseteq k[x_1, \dots, x_n]\}$ are the closed sets of a topology on C^n; that is,*

1. *$C^n = Z(\{0\})$ and $\varnothing = Z(\{1\})$.*
2. *If S and T are subsets of $k[x_1, \dots, x_n]$, then $Z(S) \cup Z(T) = Z(ST)$, where $ST = \{fg : f \in S, t \in T\}$.*
3. *If $\{S_\alpha\}$ is an arbitrary collection of subsets of $k[x_1, \dots, x_n]$, then $\bigcap_\alpha Z(S_\alpha) = Z(\bigcup_\alpha S_\alpha)$.*

Proof. The first two parts are clear from the definitions. For the third, let $P \in Z(S)$. Then $f(P) = 0$ for all $f \in S$, so $(fg)(P) = 0$ for all $fg \in ST$. Thus, $Z(S) \subseteq Z(ST)$. Similarly, $Z(T) \subseteq Z(ST)$, so $Z(S) \cup Z(T) \subseteq Z(ST)$. For the reverse inclusion, let $P \in Z(ST)$. If $P \notin Z(S)$, then there is an

$f \in S$ with $f(P) \neq 0$. If $g \in T$, then $0 = (fg)(P) = f(P)g(P)$, so $g(P) = 0$, which forces $P \in Z(T)$. Thus, $Z(ST) \subseteq Z(S) \cup Z(T)$. This proves that $Z(S) \cup Z(T) = Z(ST)$.

For the fourth part, the inclusion $Z(\bigcup_\alpha S_\alpha) \subseteq \bigcap_\alpha Z(S_\alpha)$ follows from part 1. For the reverse inclusion, take $P \in \bigcap_\alpha Z(S_\alpha)$. Then $P \in Z(S_\alpha)$ for each α, so $f(P) = 0$ for each $f \in S_\alpha$. Thus, $P \in Z(\bigcup_\alpha S_\alpha)$. □

Example 21.12 Let $GL_n(C)$ be the set of all invertible $n \times n$ matrices over C. Then $GL_n(C)$ is the complement of the zero set $Z(\det)$, so $GL_n(C)$ is an open subset of C^{n^2} with respect to the k-Zariski topology. We can view $GL_n(C)$ differently in order to view it as an algebraic variety. Let t be a new variable, and consider the zero set $Z(t\det -1)$ in C^{n^2+1}. Then the map $GL_n(C) \to Z(t\det -1)$ given by $P \mapsto (P, 1/\det(P))$ is a bijection between $GL_n(C)$ and $Z(t\det -1)$. If we introduce the definition of a morphism of varieties, this map would turn out to be an isomorphism. In Problem 10, we give the definition of a morphism between varieties.

Starting with an ideal I of $k[x_1, \ldots, x_n]$, we obtain a k-variety $Z(I)$. We can reverse this process and obtain an ideal from a k-variety.

Definition 21.13 *Let $V \subseteq C^n$. The ideal of V is*

$$I(V) = \{f \in k[x_1, \ldots, x_n] : f(P) = 0 \textit{ for all } P \in V\}.$$

The coordinate ring of V is the ring $k[V] = k[x_1, \ldots, x_n]/I(V)$.

If $f \in k[x_1, \ldots, x_n]$ and $V \subseteq C^n$, then f can be viewed as a function from V to k. Two polynomials f and g yield the same polynomial function on V if and only if $f - g \in I(V)$; hence, we see that $k[V]$ can be thought of as the ring of polynomial functions on V.

One of the main techniques of algebraic geometry is to translate back and forth from geometric properties of varieties to algebraic properties of their coordinate rings. We state Hilbert's Nullstellensatz below, the most fundamental result that connects the geometry of varieties with the algebra of polynomial rings.

Let A be a commutative ring, and let I be an ideal of A. Then the *radical* of I is the ideal

$$\sqrt{I} = \{f \in A : f^r \in I \text{ for some } r \in \mathbb{N}\}.$$

If $I = \sqrt{I}$, then I is said to be a *radical ideal.* A standard result of commutative ring theory is that $\sqrt{I}$ is the intersection of all prime ideals of A containing I (see Problem 2).

Lemma 21.14 *If V is any subset of C^n, then $I(V)$ is a radical ideal of $k[x_1, \ldots, x_n]$.*

Proof. Let $f \in k[x_1, \ldots, x_n]$ with $f^r \in I(V)$ for some r. Then $f^r(P) = 0$ for all $P \in V$. But $f^r(P) = (f(P))^r$, so $f(P) = 0$. Therefore, $f \in I(V)$; hence, $I(V)$ is equal to its radical, so $I(V)$ is a radical ideal. □

Lemma 21.15 *The following statements are some properties of ideals of subsets of C^n.*

1. *If X and Y are subsets of C^n with $X \subseteq Y$, then $I(Y) \subseteq I(X)$.*

2. *If J is a subset of $k[x_1, \ldots, x_n]$, then $J \subseteq I(Z(J))$.*

3. *If $V \subseteq C^n$, then $V \subseteq Z(I(V))$, and $V = Z(I(V))$ if and only if V is a k-variety.*

Proof. The first two parts of the lemma are clear from the definition of $I(V)$. For the third, let V be a subset of C^n. If $f \in I(V)$, then $f(P) = 0$ for all $P \in V$, so $P \in Z(I(V))$, which shows that $V \subseteq Z(I(V))$. Suppose that $V = Z(S)$ for some subset $S \in k[x_1, \ldots, x_n]$. Then $S \subseteq I(V)$, so $Z(I(V)) \subseteq Z(S) = V$ by the previous lemma. Thus, $V = Z(I(V))$. Conversely, if $V = Z(I(V))$, then V is a k-variety by definition. □

In the lemma above, if J is an ideal of $k[x_1, \ldots, x_n]$, we have $J \subseteq I(Z(J))$, and actually $\sqrt{J} \subseteq I(Z(J))$, since $I(Z(J))$ is a radical ideal. The following theorem, Hilbert's Nullstellensatz, shows that $I(Z(J))$ is always equal to $\sqrt{J}$.

Theorem 21.16 (Nullstellensatz) *Let J be an ideal of $k[x_1, \ldots, x_n]$, and let $V = Z(J)$. Then $I(V) = \sqrt{J}$.*

Proof. For a proof of the Nullstellensatz, see Atiyah and Macdonald [2, p. 85] or Kunz [19, p. 16]. □

Corollary 21.17 *There is a 1–1 inclusion reversing correspondence between the k-varieties in C^n and the radical ideals of $k[x_1, \ldots, x_n]$ given by $V \mapsto I(V)$. The inverse correspondence is given by $J \mapsto Z(J)$.*

Proof. If V is a k-variety, then the previous lemma shows that $V = Z(I(V))$. Also, the Nullstellensatz shows that if I is a radical ideal, then $J = I(Z(J))$. These two formulas tell us that the association $V \mapsto I(V)$ is a bijection and that its inverse is given by $J \mapsto Z(J)$. □

Another consequence of the Nullstellensatz is that any proper ideal defines a nonempty variety. Suppose that I is a proper ideal of $k[x_1, \ldots, x_n]$. If $V = Z(J)$, then the Nullstellensatz shows that $I(V) = \sqrt{J}$. Since J is a proper ideal, the radical $\sqrt{J}$ is also proper. However, if $Z(J) = \varnothing$, then $I(Z(J)) = k[x_1, \ldots, x_n]$. Thus, $Z(J)$ is nonempty.

Example 21.18 Let $f \in k[x_1, \ldots, x_n]$ be a polynomial, and let $V = Z(f)$. If $f = p_1^{r_1} \cdots p_t^{r_t}$ is the irreducible factorization of f, then $I(V) = \sqrt{(f)}$ by the Nullstellensatz. However, we show that $\sqrt{(f)} = (p_1 \cdots p_t)$ for, if $g \in \sqrt{(f)}$, then $g^m = fh$ for some $h \in k[x_1, \ldots, x_n]$ and some $m > 0$. Each p_i then divides g^m; hence, each p_i divides g. Thus, $g \in (p_1 \cdots p_t)$. For the reverse inclusion, $p_1 \cdots p_t \in \sqrt{(f)}$, since if r is the maximum of the r_i, then $(p_1 \cdots p_t)^r \in (f)$.

If $f \in k[x_1, \ldots, x_n]$ is irreducible, then $\sqrt{(f)} = (f)$, so the coordinate ring of $Z(f)$ is $k[x_1, \ldots, x_n]/(f)$. For example, the coordinate ring of $Z(y - x^2) \subseteq C^2$ is $k[x, y]/(y - x^2)$. This ring is isomorphic to the polynomial ring $k[t]$. Similarly, the coordinate ring of $Z(y^2 - x^3)$ is $k[x, y]/(y^2 - x^3)$. This ring is isomorphic to the subring $k[t^2, t^3]$ of the polynomial ring $k[t]$; an isomorphism is given by sending x to t^2 and y to t^3.

Definition 21.19 *Let V be a k-variety. Then V is said to be irreducible if V is not the union of two proper k-varieties.*

Every k-variety can be written as a finite union of irreducible subvarieties, as Problem 7 shows. This fact reduces many questions about varieties to the case of irreducible varieties.

Example 21.20 Let V be an irreducible k-variety. By taking complements, we see that the definition of irreducibility is equivalent to the condition that any two nonempty open sets have a nonempty intersection. Therefore, if U and U' are nonempty open subsets of V, then $U \cap U' \neq \varnothing$. One consequence of this fact is that any nonempty open subset of V is dense in V, as we now prove. If U is a nonempty open subset of V, and if C is the closure of U in V, then $U \cap (V - C) = \varnothing$. The set $V - C$ is open, so one of U or $V - C$ is empty. Since U is nonempty, this forces $V - C = \varnothing$, so $C = V$. But then the closure of U in V is all of V, so U is dense in V. This unusual fact about the Zariski topology is used often in algebraic geometry.

Proposition 21.21 *Let V be a k-variety. Then V is irreducible if and only if $I(V)$ is a prime ideal, if and only if the coordinate ring $k[V]$ is an integral domain.*

Proof. First suppose that V is irreducible. Let $f, g \in k[x_1, \ldots, x_n]$ with $fg \in I(V)$. Then $I = I(V) + (f)$ and $J = I(V) + (g)$ are ideals of $k[x_1, \ldots, x_n]$ containing $I(V)$; hence, their zero sets $Y = Z(I)$ and $Z = Z(J)$ are contained in $Z(I(V)) = V$. Moreover, $IJ \subseteq I(V)$, since $fg \in I(V)$, so $Y \cup Z = Z(IJ)$ contains V. This forces $V = Y \cup Z$, so either $Y = V$ or $Z = V$, since V is irreducible. If $Y = V$, then $I \subseteq I(Y) = I(V)$, and if $Z = V$, then $J \subseteq I(Z) = I(V)$. Thus, either $f \in I(V)$ or $g \in I(V)$, so $I(V)$ is a prime ideal of $k[x_1, \ldots, x_n]$.

Conversely, suppose that $I(V)$ is prime. If $V = Y \cup Z$ for some k-varieties Y and Z, let $I = I(Y)$ and $J = I(Z)$. Then $IJ \subseteq I(Y \cup Z) = I(V)$, so either $I \subseteq I(V)$ or $J \subseteq I(V)$. This means that $V \subseteq Z(I) = Y$ or $V \subseteq Z(J) = Z$. Therefore, $Y = V$ or $Z = V$, so V is irreducible. □

In Section 22, we will obtain finitely generated field extensions by considering the quotient field of the coordinate ring of an irreducible k-variety as an extension of k. We finish this section with a brief discussion of the dimension of a variety. In Theorem 22.5, we will see that the dimension of an irreducible variety V is equal to the transcendence degree over k of the quotient field of $k[V]$.

Definition 21.22 *Let V be a k-variety. Then the dimension of V, denoted* $\dim(V)$, *is the largest integer n such that there is a chain*

$$Y_0 \subset Y_1 \subset \cdots \subset Y_n \subseteq V$$

of irreducible k-subvarieties of V.

While it is not obvious, there is indeed a maximum among the lengths of chains of irreducible subvarieties of any variety. This is a consequence of Theorem 22.5. In fact, if $V \subseteq C^n$, then $\dim(V) \leq n$.

The definition above is purely topological. However, the dimension of a k-variety can be determined with purely algebraic methods. One way to determine the dimension of a k-variety is given in the proposition below.

Proposition 21.23 *Let V be a k-variety. Then* $\dim(V)$ *is the maximum nonnegative integer n such that there is a chain*

$$P_0 \subset P_1 \subset \cdots \subset P_n$$

of prime ideals of $k[V]$.

Proof. Suppose that $Y_0 \subset Y_1 \subset \cdots \subset Y_n \subseteq V$ is a chain of closed irreducible subsets of V. Then

$$I(V) \subseteq I(Y_n) \subset \cdots \subset I(Y_0)$$

is a chain of prime ideals of $k[x_1, \ldots, x_n]$ by the previous proposition. Moreover, the inclusions are proper by the Nullstellensatz. By taking images in the quotient ring $k[V] = k[x_1, \ldots, x_n]/I(V)$, we get a chain of prime ideals of length n. However, if we have a chain of prime ideals of $k[V]$ of length n, then we get a chain $I(V) \subseteq Q_0 \subset Q_1 \subset \cdots \subset Q_n$ of prime ideals of $k[x_1, \ldots, x_n]$. Taking zero sets gives a chain

$$Z(Q_n) \subset \cdots \subset Z(Q_0) \subseteq Z(I(V)) = V$$

of irreducible k-subvarieties in V. The maximum length of a chain of irreducible k-subvarieties of V is then the maximum length of a chain of prime ideals of $k[V]$. □

If A is a commutative ring, then the supremum of integers n such that there is a chain of prime ideals of A of length n is called the *dimension* of A. The proposition says that $\dim(V) = \dim(k[V])$ if V is a k-variety. Calculating the dimension of a k-variety by either the definition or by use of the proposition above is not easy. Instead, we will use Theorem 22.5 to calculate the dimension of a variety.

Problems

1. Let k be an infinite field. If $f \in k[x_1, \ldots, x_n]$ with $f(P) = 0$ for all $P \in k^n$, show that $f = 0$ in $k[x_1, \ldots, x_n]$.

2. Let A be a commutative ring, and let I be an ideal of A. Show that $\sqrt{I}$ is the intersection of all prime ideals of A containing I.
(Hint: One inclusion is easy. For the other inclusion, show that if S is a multiplicatively closed subset of A, and P is an ideal maximal among all ideals J with $J \cap S = \varnothing$, then P is prime.)

3. Let W be a subset of C^n. If $\overline{W}$ is the closure of W in the k-Zariski topology on C^n, show that $\overline{W} = Z(I(W))$.

4. Use the Nullstellensatz to show that if C is algebraically closed, then every maximal ideal of $C[x_1, \ldots, x_n]$ is of the form $(x_1 - a_1, \ldots, x_n - a_n)$ for some $a_i \in C$.

5. A topological space V is said to be *Noetherian* if V satisfies the accending chain condition (ACC) on open subsets: If $U_1 \subseteq U_2 \subseteq \cdots$ is an increasing chain of open subsets of V, then there is an n with $U_n = U_{n+r}$ for each $r \geq 0$. Show that the following statements are equivalent:

   (a) The space V is a Noetherian space.

   (b) Any nonempty collection $\{U_\alpha\}$ of open subsets of V has a maximal element; that is, there is a $U \in \{U_\alpha\}$ not properly contained in any other element of $\{U_\alpha\}$.

   (c) The space V satisfies the descending chain condition (DCC) on closed sets: If $C_1 \supseteq C_2 \cdots$ is a decreasing chain of closed subsets of V, then there is an n with $C_n = C_{n+r}$ for each $r \geq 1$.

6. Let V be a topological space. Show that V is a Noetherian space if and only if every open subset of V is compact.

7. Let V be a Noetherian topological space.

(a) Show that V can be written as a finite union of closed irreducible subsets.

(b) Suppose that $V = Y_1 \cup \cdots \cup Y_n$ with each Y_i a closed irreducible subset of V. If $Y_i \not\subseteq Y_j$ for each $i \neq j$, show that the Y_i are uniquely determined by this decomposition for V.

8. The *Hilbert basis theorem* says that $k[x_1, \ldots, x_n]$ is a Noetherian ring; that is, $k[x_1, \ldots, x_n]$ satisfies the ACC on ideals.

 (a) Show that the Hilbert basis theorem implies that C^n is a Noetherian space.

 (b) Show that the quotient ring of a Noetherian ring is also Noetherian, and conclude that any k-variety is a Noetherian space.

9. Let V be a k-variety of dimension 1. Show that any proper closed subset of V is finite.
(Hint: Show that any proper closed irreducible subset of V is a single point. Use the previous problems that show that a k-variety can be decomposed into closed irreducible subsets.)

10. Let $V \subseteq C^n$ and $W \subseteq C^m$ be k-varieties. If $f_1, \ldots, f_m \in k[x_1, \ldots, x_n]$, then the map $\varphi : V \to W$ defined by $\varphi(P) = (f_1(P), \ldots, f_m(P))$ is called a *k-morphism*. A k-isomorphism from V to W is a k-morphism whose inverse function is also a k-morphism. Two varieties are said to be k-isomorphic if there is a k-isomorphism from one to the other.

 (a) If $\varphi : V \to W$ is a k-morphism, show that there is a k-algebra map $\varphi^* : k[W] \to k[V]$ induced by sending y_i to f_i, if $k[W] = k[y_1, \ldots, y_m]/I(W)$.

 (b) Conversely, suppose that $\tau : k[W] \to k[V]$ is a k-algebra map. Use τ to define a k-morphism τ' from V to W.

 (c) If $\varphi : V \to W$ is a k-morphism, show that $(\varphi^*)' = \varphi$, and if $\tau : k[W] \to k[V]$ is a k-algebra map, show that $(\tau')^* = \tau$.

 (d) Conclude that V and W are k-isomorphic if and only if $k[V]$ and $k[W]$ are isomorphic as k-algebras.

 (A k-algebra map is a ring homomorphism that is simultaneously a k-vector space homomorphism.)

11. Show that a morphism between two k-varieties is a continuous map relative to the k-Zariski topology.

12. Let V be the k-variety C^1, and let W be the k-variety $Z(y^2 - x^3)$. Show that the map $\varphi : V \to W$ given by $\varphi(t) = (t^2, t^3)$ is a k-morphism that is a bijection and where the inverse function φ^{-1} is continuous but that φ is not a k-isomorphism.

22 Algebraic Function Fields

In this section, we study one of the most important classes of field extensions, those arising from algebraic geometry. We will continue to use the notation defined in Section 21. The point of this section is to show how field theoretic information can be used to obtain geometric information about varieties.

Definition 22.1 *Let V be an irreducible k-variety. Then the function field $k(V)$ of V is the quotient field of the coordinate ring $k[V]$.*

This definition is meaningful because if V is irreducible, then $I(V)$ is a prime ideal, so $k[V] = k[x_1, \dots, x_n]/I(V)$ is an integral domain. The function field $k(V)$ of a variety V can be viewed as a field of functions on V in the following way. Each $f \in k[V]$ is a polynomial function from V to C. A quotient f/g of elements of $k[V]$ then defines a function from $V - Z(g)$ to C. Now, $V - Z(g)$ is an open subset of V; hence, it is a dense subset of V. The elements of $k(V)$ are then rational functions defined on an open, dense subset of V; the density follows by Example 21.20.

Example 22.2 Let $V = Z(y - x^2)$. Then the coordinate ring of V is $k[x, y]/(y - x^2)$, which is isomorphic to the polynomial ring $k[t]$ by sending t to the coset of x in $k[V]$. Therefore, the function field of V is the rational function field $k(t)$.

Example 22.3 Let $V = Z(y^2 - x^3)$. Then $k(V)$ is the field $k(s, t)$, where s and t are the images of x and y in $k[V] = k[x, y]/(y^2 - x^3)$, respectively. Note that $t^2 = s^3$. Let $z = t/s$. Substituting this equation into $t^2 = s^3$ and simplifying shows that $s = z^2$, and so $t = z^3$. Thus, $k(V) = k(z)$. The element z is transcendental over k, since if $k(V)/k$ is algebraic, then $k[V]$ is a field by the argument in Example 19.11, so $(y^2 - x^3)$ is a maximal ideal of $k[x, y]$. However, this is not true, since $(y^2 - x^3)$ is properly contained in the ideal (x, y). Thus, $k(V)$ is a rational function field in one variable over k. Note that $k[V]$ is isomorphic to $k[x^2, x^3]$, a ring that is not isomorphic to a polynomial ring in one variable over k.

Example 22.4 If V is an irreducible k-variety, then V gives rise to a field extension $k(V)$ of k. We can reverse this construction. Let K be a finitely generated field extension of k. Say $K = k(a_1, \dots, a_n)$ for some $a_i \in K$. Let

$$P = \{f \in k[x_1, \dots, x_n] : f(a_1, \dots, a_n) = 0\}.$$

Then P is the kernel of the ring homomorphism $\varphi : k[x_1, \dots, x_n] \to K$ that sends x_i to a_i, so P is a prime ideal. If $V = Z(P)$, then V is an irreducible k-variety with coordinate ring $k[x_1, \dots, x_n]/P \cong k[a_1, \dots, a_n]$,

so the function field of V is K. Note that if we start with an irreducible k-variety V and let $K = k(V)$, then the variety we get from this construction may not be V. Therefore, the processes of obtaining field extensions from varieties and vice versa are not inverses of each other.

The next theorem gives the most useful method for computing the dimension of a variety. We do not give the proof, since this would go past the interests of this book. The interested reader can find a proof in Kunz [19, §3, Prop. 3.11].

Theorem 22.5 *Let V be an irreducible k-variety. Then the dimension of V is equal to the transcendence degree of $k(V)/k$.*

Example 22.6 The dimension of the k-variety C^n is n, since the function field of C^n is $k(x_1, \ldots, x_n)$, which has transcendence degree n over k.

Example 22.7 If $V = Z(y - x^2)$, then $k[V] = k[x,y]/(y - x^2) \cong k[x]$, so $k(V) \cong k(x)$ has transcendence degree 1 over k. Thus, $\dim(V) = 1$. More generally, if $f(x,y)$ is any irreducible polynomial in $k[x,y]$ and $V = Z(f)$, then $k[V] = k[x,y]/(f) = k[s,t]$, where s and t are the images in $k[V]$ of x and y, respectively. Therefore, $k(V) = k(s,t)$. The set $\{s,t\}$ is algebraically dependent over k, since $f(s,t) = 0$. However, s or t is transcendental over k, for if s is algebraic over k, then there is a $g \in k[x]$ with $g(s) = 0$. Viewing $g(x)$ as a polynomial in x and y, we see that $g \in I(V) = (f)$. Similarly, if t is algebraic over k, then there is an $h(y) \in k[y]$ with $h \in (f)$. These two inclusions are impossible, since $g(x)$ and $h(y)$ are relatively prime. This proves that either $\{s\}$ or $\{t\}$ is a transcendence basis for $k(V)$, so $k(V)$ has transcendence degree 1 over k.

Example 22.8 Let $f \in k[x_1, \ldots, x_n]$ be an irreducible polynomial and set $V = Z(f)$. Then $\dim(V) = n - 1$. To see this, we showed in Example 19.12 that the quotient field of $k[x_1, \ldots, x_n]/(f)$ has transcendence degree $n - 1$ over k. But, this quotient field is the function field $k(V)$ of V. Thus, Theorem 22.5 shows that $\dim(V) = n - 1$. Note that the argument in the previous example is mostly a repeat of that given in Example 19.12 in the case of two variables.

We now give some properties of the function field of an irreducible variety. We first need two definitions. If K/k is a field extension, then K is a *regular extension* of k provided that K/k is separable and k is algebraically closed in K. If P is a prime ideal of $k[x_1, \ldots, x_n]$, then P is *absolutely prime* if for any field extension L/k the ideal generated by P in $L[x_1, \ldots, x_n]$ is a prime ideal.

Example 22.9 Let P be an absolutely prime ideal of $k[x_1, \ldots, x_n]$, and let $V = Z(P)$. Let L be any field extension of k contained in C. Then we can

view V as an L-variety. The coordinate ring of V considered as an L-variety is $L[x_1, \ldots, x_n]/I$, where I is the ideal of V computed in $L[x_1, \ldots, x_n]$. The ideal I contains P, so I contains the ideal generated by P in $L[x_1, \ldots, x_n]$. Since P is absolutely prime, the Nullstellensatz tells us that I is the ideal generated by P. Consequently, V is irreducible as an L-variety.

If $k = \mathbb{R}$ and $P = (x^2 + y^2) \in \mathbb{R}[x, y]$, then $V = Z(P)$ is an irreducible $\mathbb{R}$-variety but V is not irreducible as a $\mathbb{C}$-variety, since the ideal of V in $\mathbb{C}[x, y]$ is $(x^2 + y^2) = (x + iy)(x - iy)$.

Theorem 22.10 *Let V be an irreducible k-variety. Then $k(V)$ is a finitely generated extension of k. Moreover, $k(V)/k$ is a regular extension if $I(V)$ is absolutely prime.*

Proof. The field $k(V)$ is the quotient field of $k[V] = k[x_1, \ldots, x_n]/I(V)$. The ring $k[V]$ is generated over k as a ring by the images of the x_i, so $k(V)$ is generated as a field extension over k by the images of the x_i. This proves that $k(V)$ is a finitely generated extension of k.

Suppose that $I(V)$ is absolutely prime. We need to show that $k(V)/k$ is separable and that k is algebraically closed in $k(V)$. For this, we first show that if L is any extension of k, then $k(V)$ and L are linearly disjoint over k. To see this, note that

$$k[V] \otimes_k L \cong L[x_1, \ldots, x_n]/Q,$$

where $Q = I(V)L[x_1, \ldots, x_n]$. This isomorphism is given on generators by $(f + I(V)) \otimes l \mapsto fl + Q$. The ring $L[x_1, \ldots, x_n]/Q$ contains an isomorphic copy of $k[V] = k[x_1, \ldots, x_n]/I(V)$, and it is the ring generated by L and this copy of $k[V]$. By the assumption that $I(V)$ is absolutely prime, Q is a prime ideal, so $L[x_1, \ldots, x_n]/Q$ is a domain. If K is the quotient field of this domain, there are isomorphic copies of $k[V]$ and L inside K, and the tensor product $k[V] \otimes_k L$ is isomorphic to a subring of K. Therefore, $k[V]$ and L are linearly disjoint over k, so $k(V)$ and L are linearly disjoint over k by Lemma 20.10. To see that $k(V)$ is separable over k, set $L = k^{1/p^\infty}$. From what we have shown, $k(V)$ and k^{1/p^∞} are linearly disjoint, so $k(V)$ is separable over k. Let k' be the algebraic closure of k in $k(V)$. By setting $L = k'$, since $k(V)$ and k' are linearly disjoint over k, it follows that k' and k' are linearly disjoint over k, so $k' = k$. Thus, k is algebraically closed in $k(V)$. This finishes the proof that $k(V)$ is a regular extension of k. □

Corollary 22.11 *Let $f \in k[x_1, \ldots, x_n]$ be an absolutely irreducible polynomial. If $V = Z(f)$, then V is an irreducible k-variety, and $k(V)$ is a regular extension of k.*

Proof. Since f is irreducible in $k[x_1, \ldots, x_n]$, the principal ideal (f) is prime; hence, $I(V) = (f)$ is prime. Thus, V is an irreducible k-variety. Moreover, (f) is absolutely prime, since f is absolutely irreducible. By the previous theorem, $k(V)$ is a regular extension of k. □

Example 22.12 Let $f = y^2 - (x^3 - x)$ and $V = Z(f)$. If L/k is any field extension, then f is irreducible in $L[x, y]$, since $x^3 - x$ is not a square in $L[x]$. Therefore, $k(V)$ is a regular extension of k.

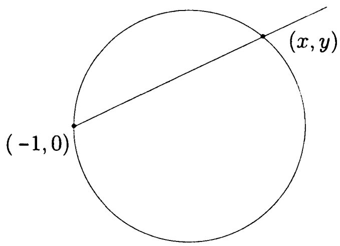
Example 22.13 If $f = x^2 + y^2 \in \mathbb{R}[x, y]$ and $V = Z(f)$, then f is irreducible over $\mathbb{R}$, but f is not irreducible over $\mathbb{C}$, since $f = (x + iy)(x - iy)$. The field extension $\mathbb{R}(V)/\mathbb{R}$ is therefore not regular. This extension is separable, since $\text{char}(\mathbb{R}) = 0$. In $\mathbb{R}(V)$, we have $x^2 + y^2 = 0$, so $(x/y)^2 = -1$. Thus, $\mathbb{C}$ is a subfield of $\mathbb{R}(V)$, which shows that $\mathbb{R}$ is not algebraically closed in $\mathbb{R}(V)$.

A natural question to ask is what geometric information about a variety can be determined from field theoretic information about its function field. Problem 6 below investigates one aspect of this question. We now investigate another.

Definition 22.14 *An irreducible k-variety V is said to be rational if $k(V)$ is a purely transcendental extension of k.*

Recall that a purely transcendental extension with finite transcendence degree is often called a rational extension. Thus, a k-variety V is rational if $k(V)/k$ is a rational extension. A fundamental problem of algebraic geometry is to determine when a variety is rational. The problem of rationality has a more geometric formulation. Recall from vector calculus that a curve in $\mathbb{R}^2$ can be parameterized in the form $x = f(t)$ and $y = g(t)$, where f and g are real-valued functions; that is, the curve consists of the points $(f(t), g(t))$ as t ranges over $\mathbb{R}$. The functions f and g can be completely general, and even with a curve defined by polynomial equations, the functions f and g may be transcendental. For example, the most common parameterization of the unit circle is $x = \cos t$ and $y = \sin t$. In the case of algebraic varieties, we are interested in parameterizations involving polynomial or rational functions.

Example 22.15 Let V be the zero set of $x^2 + y^2 - 1$, an irreducible k-variety in C^2. As noted above, if $k = \mathbb{R}$, then the curve V has a transcendental parameterization. We wish to find a parameterization of V in terms of rational functions. We can do this as follows.



Pick a point on V, for instance $P = (-1, 0)$. For a point (x, y) on V, let t be the slope of the line connecting these two points. Then $t = y/(x+1)$. If we solve for y and substitute into the equation $x^2 + y^2 - 1 = 0$, we can solve for x in terms of t. Doing this, we see that

$$x = \frac{1-t^2}{1+t^2}, \quad y = \frac{2t}{1+t^2}.$$

Moreover, we can reverse this calculation to show that

$$\left\{ \left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2} \right) : t \in C, t^2 \neq -1 \right\} = V - \{P\},$$

for, given $(x, y) \in V$ with $(x, y) \neq (-1, 0)$, solving for t in the equation

$$(1-t^2)/(1+t^2) = x$$

yields

$$t = \pm\sqrt{\frac{1-x}{1+x}},$$

which are elements of C, since $1 + x \neq 0$ and C is algebraically closed, so C contains a square root of any element. With either of these values of t, we see that $2t/(1+t^2) = t(1+x)$, and we can check that $x^2 + (t(1+x))^2 = 1$; hence, $y = 2t/(1+t^2)$ if the sign of the square root is chosen appropriately. So, this parameterization of V picks up all but one point of V. There is no value of t that yields the point P. Intuitively, we would need $t = \infty$ to get $x = -1$ and $y = 0$. Starting with any point Q on the curve and following this procedure will yield a parameterization of $V - \{Q\}$.

Example 22.16 For another example of a parameterization, let $Y = Z(y^2 - x^3)$. If we start with the point $(0, 0)$ and follow the procedure of Example 22.15, we obtain the parameterization $x = t^2$ and $y = t^3$ given in Example 21.6. With this parameterization, we get all points of Y; that is,

$$Y = \left\{ (t^2, t^3) : t \in C \right\}.$$

Not every algebraic curve can be parameterized with rational functions. To give an intuitive feel for why this is true, let V be the zero set of $y^2 - (x^3 - x)$. Pick $P = (0, 0)$ on V. If we follow the procedure above, we would get $t = y/x$, or $y = tx$. Substituting this into the equation $y^2 = x^3 - x$ yields $t^2x^2 = x^3 - x$, or $x^2 - t^2x - 1 = 0$. This has the two solutions

$$x = \frac{t^2 \pm \sqrt{t^2+4}}{2},$$

neither of which are rational functions in t. While this does not prove that Y cannot be parameterized, it does indicate that Y is more complicated

than the two previous examples. In Proposition 22.18, we show that an irreducible curve V can be parameterized if and only if the function field $k(V)$ is rational over k. A proof that $\mathbb{C}(V)/\mathbb{C}$ is not rational if $V = Z(y^2 - x^3 + x)$ is outlined in Problem 23.6. It is nontrivial to show that a field extension K/F is not rational when F is algebraically closed. If F is not algebraically closed, then it is easier to prove that an extension of F is not rational, as can be seen in Problems 1 and 4.

We now relate the concept of parameterization to that of rationality. We make precise what it means to parameterize a variety. We will restrict to curves. An algebraic variety of dimension 1 is said to be a *curve*.

Definition 22.17 *Let $V \subseteq C^n$ be a curve defined over k. Then V can be parameterized if there are rational functions $f_i(t) \in k(t)$ such that $\{(f_1(t), \ldots, f_n(t)) : t \in C\}$ is a dense subset of V with respect to the k-Zariski topology.*

From Theorem 22.5, the function field of a curve defined over a field k has transcendence degree 1 over k. We could define what it means to parameterize a variety of dimension greater than 1, although we will not do so.

To clarify the definition above, if $f(t)$ is a rational function, say $f(t) = g(t)/h(t)$ with $g, h \in k[t]$. Then $f(a)$ is defined for $a \in C$ only if $h(a) \neq 0$. The polynomial h has at most finitely many roots, so $f(a)$ is defined at all but finitely many $a \in C$. In the definition of parameterization of a curve, it is being assumed that the point $(f_1(t), \ldots, f_n(t))$ exists only when each $f_i(t)$ is defined.

Proposition 22.18 *Let V be an irreducible curve defined over k. Then V can be parameterized if and only if the function field $k(V)$ is rational over k.*

Proof. First, suppose that $V \subseteq C^n$ can be parameterized. Let $f_1(t), \ldots, f_n(t) \in k(t)$ such that $U = \{(f_1(t), \ldots, f_n(t)) : t \in C\}$ is a dense subset of V. Define $\varphi : k[x_1, \ldots, x_n] \to k(t)$ by sending x_i to $f_i(t)$. Then φ uniquely defines a k-homomorphism. The kernel of φ consists of all polynomials $h(x_1, \ldots, x_n)$ with $h(f_1(t), \ldots, f_n(t)) = 0$. For such an h, we have $h(P) = 0$ for all $P \in U$. Therefore, $U \subseteq Z(h)$, so by density we have $V \subseteq Z(h)$. Thus, $h \in I(V)$. It is clear that $I(V) \subseteq \ker(\varphi)$; hence, we see that $\ker(\varphi) = I(V)$, so φ induces an injective k-homomorphism $\varphi' : k[V] \to k(t)$. The map φ' then induces a k-homomorphism $k(V) \to k(t)$, so $k(V)$ is isomorphic to an intermediate field of $k(t)/k$. By Lüroth's theorem, which we prove below, $k(V)$ is a rational extension of k.

For the converse, suppose that $k(V) = k(t)$ for some t. We abuse notation by writing x_i for the image of x_i in $k[V]$. We have $x_i = f_i(t)$ for some rational function f_i, and we can write $t = g(x_1, \ldots, x_n)/h(x_1, \ldots, x_n)$ for some polynomials g, h. If $P \in V$, let $a = g(P)/h(P)$, provided that $h(P) \neq$

0. Then $P = (f_1(a), \dots, f_n(a))$ by the relations between the x_i and t. On the other hand, given $a \in C$, if each $f_i(a)$ is defined, let $Q = (f_1(a), \dots, f_n(a))$. Then $u(Q) = 0$ for all $u \in I(V)$, again by the relations between the x_i and t. Thus, $Q \in Z(I(V)) = V$. The points of V not of the form $(f_1(a), \dots, f_n(a))$ all satisfy $h(P) = 0$. This does not include all points of V, or else $h \in I(V)$, which is false by the choice of h. Thus, $V \cap Z(h)$ is a finite set, so $\{(f_1(t), \dots, f_n(t)) : t \in C\}$ contains all but finitely many points of V, so it is a dense subset of V. The equations $x_i = f_i(t)$ thus give a parameterization of V. □

We now finish the proof of Proposition 22.18 by proving Lüroth's theorem.

Theorem 22.19 (Lüroth) *Let $k(t)$ be the rational function field in one variable over a field k, and let F be a field with $k \subset F \subseteq k(t)$. Then $F = k(u)$ for some $u \in F$. Thus, F is purely transcendental over k.*

Proof. Let $K = k(t)$, and take $v \in F - k$. We have seen in Example 1.17 that $[K : k(v)] < \infty$, so $[K : F] < \infty$. Let $f(x) = x^n + l_{n-1}x^{n-1} + \cdots + l_0$ be the minimal polynomial of t over F. Then $[K : F] = n$. Since t is transcendental over k, some $l_i \notin k$. Let $u = l_i$, and set $m = [K : k(u)]$. Therefore, $m \geq n$, since $k(u) \subseteq F$. If we show $m \leq n$, then we will have proved that $F = k(u)$. All $l_j \in k(t)$, so there are polynomials $c_1(t), \dots, c_n(t)$ and $d(t)$ in $k[t]$ with $l_j = c_j(t)/d(t)$, and such that $\{d, c_1, \dots, c_n\}$ is relatively prime. Note that $c_n(t) = d(t)$, since f is monic, and $u = c_i(t)/d(t)$, so $m \leq \max\{\deg(c_i), \deg(d)\}$ by Example 1.17. This may be an inequality instead of an equality because c_i and d may not be relatively prime. Let

$$f(x,t) = d(t)f(x) = c_n(t)x^n + c_{n-1}(t)x^{n-1} + \cdots + c_0(t).$$

Then $f(x,t) \in k[x,t]$, and f is primitive as a polynomial in x. Moreover, $\deg_x(f(x,t)) = n$, where $\deg_x$ refers to the degree in x of a polynomial, and $\deg_t(f(x,t)) \geq m$, since c_i and d are both coefficients of f. By dividing out $\gcd(c_i, d)$, we may write $u = g(t)/h(t)$ with $g, h \in k[t]$ relatively prime. Now t is a root of $g(x) - uh(x) \in F[x]$, so we may write

$$g(x) - uh(x) = q(x)f(x) \tag{22.1}$$

with $q(x) \in F[x]$. Plugging $u = g(t)/h(t)$ into Equation (22.1), we see that $g(x)h(t) - g(t)h(x)$ is divisible by $f(x,t)$ in $k(t)[x]$ as $F \subseteq k(t)$. These polynomials are in $k[x,t]$, and f is primitive in x, so we can write

$$g(x)h(t) - g(t)h(x) = r(x,t)f(x,t)$$

with $r(x,t) \in k[x,t]$. The left-hand side has degree in t at most m, since $m = \max\{\deg(g), \deg(h)\}$; this equality was proved in Example 1.17. But

we know that the degree of f in t is at least m. Thus, $r(x,t) = r(x) \in k[x]$. In particular, r is primitive as a polynomial in $k[t][x]$. Thus, rf is primitive in $k[t][x]$ by Proposition 4.3 of Appendix A, so $l(x,t) = g(x)h(t) - g(t)h(x)$ is a primitive polynomial in $k[t][x]$. By symmetry, it is also primitive in $k[x][t]$. But $r(x)$ divides all of its coefficients, so $r \in k$. Thus,

$$\begin{aligned} n &= \deg_x(f) = \deg_x(g(x)h(t) - g(t)h(x)) \\ &= \deg_t(g(x)h(t) - g(t)h(x)) \\ &= \deg_t(f) \geq m. \end{aligned}$$

Therefore, $n \geq m$. Since we have already proved that $n \leq m$, we get $n = m$, and so $F = k(u)$. □

Lüroth proved this theorem in 1876. It led to the following rationality problem: If L is an intermediate field of $k(x_1, \ldots, x_n)/k$ with $\mathrm{trdeg}(L/k) = n$, is L/k rational? Castelnuovo proved in 1893 that this is true for $n = 2$ if k is algebraically closed. It was not until the early 1970s, however, that an example of an intermediate field of $\mathbb{C}(x,y,z)/\mathbb{C}$ that is not rational over $\mathbb{C}$ was found.

Problems

1. Let $V \subseteq \mathbb{C}^2$ be the zero set of $x^2 + y^2 + 1 = 0$. Then V is defined over $\mathbb{R}$ and over $\mathbb{C}$. Show that the function field of V is isomorphic to $k(t)(\sqrt{-1-t^2})$, where $k = \mathbb{R}$ or $k = \mathbb{C}$, depending on what we take to be the base field. Show that $\mathbb{R}(V)$ is not rational over $\mathbb{R}$ but that $\mathbb{C}(V)$ is rational over $\mathbb{C}$.

2. Let V be as in Problem 1. Find a parameterization of V over $\mathbb{C}$.

3. Let V be as in Problem 1. By Proposition 22.18, there is no parameterization of V over $\mathbb{R}$. Verify this directly for V by showing that the set of $\mathbb{R}$-rational points is nonempty and then showing that a parameterized curve always has rational points.

4. Let k be a field of characteristic not 2, and let $a, b \in k^*$. Show that $ax^2 + by^2 - 1$ is irreducible over k. Let K be the function field of the k-variety $V = Z(ax^2 + by^2 - 1)$. Show that K/k is rational if and only if V has a k-rational point.
(Note: Problem 1 can be viewed as a special case of this.)

5. Let V be the zero set of $\{z^2 - x^2y, y^2 - xz\}$. Mimic the method of Example 22.15 to find a way to parameterize curves in C^3, and starting with the point $P = (0,0,0)$, obtain the parameterization (t^3, t^4, t^5) of V.

6. Let V and W be irreducible k-varieties. A *rational map* from V to W is a map $\varphi : U \to W$ defined on a dense open subset U of V, of the form $\varphi(P) = (f_1(P), \dots, f_m(P))$ for some $f_i \in k(x_1, \dots, x_n)$. We will write $\varphi : V \to W$ even though the domain of φ may be a proper subset of V. A rational map φ is said to be *dominant* if $\text{im}(\varphi)$ is dense in W. If $\tau : k(W) \to k(V)$ is a k-homomorphism of fields, show that τ induces a dominant rational map $V \to W$.
(While we have not defined k-morphism except on closed subsets of C^n, if we extend the definition in an appropriate way, a consequence of this problem is that $k(V)$ and $k(W)$ are k- isomorphic field extensions of k if and only if there are dense, open subsets $V_0 \subseteq V$ and $W_0 \subseteq W$ such that V_0 and W_0 are k-isomorphic.)

7. Let V and W be irreducible k-varieties. Then V and W are said to be *birational*, provided that there are rational maps $\varphi : V \to W$ and $\phi : W \to V$ such that $\varphi \circ \phi$ and $\phi \circ \varphi$ are each the identity on their respective domains. Show that V and W are birational if and only if their function fields are k-isomorphic.

8. Let V be an irreducible k-variety, and assume that k is perfect. Show that V is birational to a *hypersurface* $Z(f)$ for some $f \in k[x_1, \dots, x_n]$.

9. Let V be a k-variety, and suppose that $I(V)$ is an absolutely prime ideal. If L is an extension field of k contained in C, show that the function field $L(V)$ of V viewed as an L-variety is the free join of L and $k(V)$.
(See the definition in Problem 20.7.)

10. Let V be an irreducible k-variety. Assume that C is an algebraic closure of k, and let $G = \text{Gal}(C/k)$. For $\sigma \in G$, let σ act on C^n by

$$\sigma((a_1, \dots, a_n)) = (\sigma(a_1), \dots, \sigma(a_n)).$$

(a) Show that $\sigma \in G$ sends V to V.

(b) Show that $\sigma \in G$ induces a homomorphism from $C[x_1, \dots, x_n]$ to itself and fixes $k[x_1, \dots, x_n]$.

(c) Let $J = I(V)C[x_1, \dots, x_n]$. Assume that

$$J = \{f \in C[x_1, \dots, x_n] : f(P) = 0 \text{ for all } P \in V\}.$$

Show that $\sigma \in G$ sends J to itself and, hence, σ induces a homomorphism from $C[x_1, \dots, x_n]/J$ to itself, and that the subring of $C[x_1, \dots, x_n]/J$ that is fixed by G is $k[V]$.
(The assumption above can be shown to hold if k is a perfect field; see Problem 11.)

11. Read §6 of Draxl [6] and use Theorem 1 of [6] to prove the following statement: If k is perfect, then

$$k(V)C[x_1, \dots, x_n] = \{f \in C[x_1, \dots, x_n] : f(P) = 0 \text{ for all } P \in V\}.$$

In other words, viewing V as a C-variety, the ideal of V is generated by polynomials defined over k.

23 Derivations and Differentials

In this section, we discuss algebraic notions of derivation and differential, and we use these concepts to continue our study of finitely generated field extensions. We shall see that by using differentials we can determine the transcendence degree of a finitely generated extension and when a subset of a separably generated extension is a separating transcendence basis. As a geometric application, we use these ideas to define the tangent space to a point of a variety. By using tangent spaces, we are able to define the notion of nonsingular point on a variety. This is a more subtle geometric concept than those discussed in Section 21.

Let A be a commutative ring, and let M be an A-module. A *derivation of* A into M is a map $D : A \to M$ such that for all $a, b \in A$,

$$\begin{aligned} D(a+b) &= D(a) + D(b), \\ D(ab) &= bD(a) + aD(b). \end{aligned}$$

We write $\operatorname{Der}(A, M)$ for the set of all derivations of A into M. Since the sum of derivations is easily seen to be a derivation, $\operatorname{Der}(A, M)$ is a group. Furthermore, $\operatorname{Der}(A, M)$ is an A-module by defining $aD : A \to M$ by $(aD)(x) = a(D(x))$.

Example 23.1 The simplest example of a derivation is the polynomial derivative map $d/dx : k[x] \to k[x]$ defined by

$$\frac{d}{dx}\left(\sum_{i=0}^{n} a_i x^i\right) = \sum_{i=1}^{n-1} i a_i x^{i-1},$$

where k is any commutative ring. The term ia_i in the formula above is, of course, the sum of a_i with itself i times.

Example 23.2 If k is a field, then the derivation d/dx on $k[x]$ can be extended to the quotient field $k(x)$ by use of the quotient rule; that is, the formula

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{g(x)\frac{d}{dx}f(x) - f(x)\frac{d}{dx}g(x)}{g(x)^2}$$

defines a derivation on $k(x)$. We shall see a generalization of this example in Lemma 23.10.

Example 23.3 Let k be any commutative ring, and let $A = k[x_1, \dots, x_n]$ be the polynomial ring in n variables over k. Then the partial derivative maps $\partial/\partial x_i$ are each derivations of A to itself.

Example 23.4 Let K be a field, and let $D \in \mathrm{Der}(K, K)$. If $a \in K^*$, we prove that $D(a^{-1}) = -a^{-2}D(a)$. To see this, note that $D(1) = 0$ by an application of the product rule. Thus,

$$\begin{aligned} 0 = D(1) &= D(a \cdot a^{-1}) \\ &= a^{-1}(D(a)) + aD(a^{-1}). \end{aligned}$$

Solving for $D(a^{-1})$ gives $D(a^{-1}) = -a^{-2}D(a)$, as desired.

Other familiar facts from calculus can be verified for arbitrary derivations. For instance, if K is a field and $a, b \in K$ with $b \neq 0$, and if $D \in \mathrm{Der}(K, K)$, then

$$D\left(\frac{a}{b}\right) = \frac{bD(a) - aD(b)}{b^2}.$$

To see this, we have

$$\begin{aligned} D(ab^{-1}) &= b^{-1}D(a) + aD(b^{-1}) \\ &= b^{-1}D(a) - ab^{-2}D(b) \\ &= b^{-2}(bD(a) - aD(b)) \end{aligned}$$

from the previous calculation. This proves the validity of the quotient rule for derivations on a field.

Let D be a derivation of a ring A into an A-module M. An element $a \in A$ is said to be a *constant* for D if $D(a) = 0$. It is not hard to see that the set of all constants for D is a subring of A. If B is a subring of A, let $\mathrm{Der}_B(A, M)$ be the set of all derivations $D : A \to M$ for which $D(b) = 0$ for all $b \in B$. By studying $\mathrm{Der}_B(A, A)$, we will obtain information about the extension A/B when A and B are fields. To simplify notation, let $\mathrm{Der}_B(A) = \mathrm{Der}_B(A, A)$. We will call an element of $\mathrm{Der}_B(A)$ a B-derivation on A.

Let K be a field extension of F. We wish to see how the vector space $\mathrm{Der}_F(K)$ gives information about the field extension K/F, and vice versa. We first consider algebraic extensions. The following lemma, which can be thought of as the chain rule for derivations, will be convenient in a number of places.

Lemma 23.5 *Let K be a field extension of k, and let $D \in \mathrm{Der}_k(K)$. If $a \in K$ and $f(x) \in k[x]$, then $D(f(a)) = f'(a)D(a)$, where $f'(x)$ is the ordinary polynomial derivative of f. More generally, if $f(x_1, \dots, x_n) \in k[x_1, \dots, x_n]$*

and $a_1, \ldots, a_n \in K$, *then*

$$D(f(a_1, \ldots, a_n)) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a_1, \ldots, a_n) D(a_i).$$

Proof. Suppose that $f(x) = \sum \alpha_i x^i$. Then

$$\begin{aligned} D(f(a)) &= D\left(\sum \alpha_i a^i\right) \\ &= \sum \alpha_i D(a^i) = \sum \alpha_i i a^{i-1} D(a) \\ &= f'(a) D(a). \end{aligned}$$

The second statement follows from much the same calculation. If $f = \sum_{\mathbf{i}} \alpha_{\mathbf{i}} x_1^{i_1} \cdots x_n^{i_n}$, where $\mathbf{i} = (i_1, \ldots, i_n)$, applying the property $D(ab) = bD(a) + aD(b)$ repeatedly, we see that

$$\begin{aligned} D(f(a_1, \ldots, a_n) &= \sum_{j=1}^{n} \sum_{\mathbf{i}} i_j \alpha_{\mathbf{i}} a_1^{i_1} \cdots a_{j-1}^{i_{j-1}} a_j^{i_j - 1} D(a_j) a_{j+1}^{i_{j+1}} \cdots a_n^{i_n} \\ &= \sum_{j=1}^{n} \frac{\partial f}{\partial x_i}(a_1, \ldots, a_n) D(a_j). \end{aligned}$$

□

Proposition 23.6 *Let* K *be a separable algebraic field extension of* F. *Then* $\mathrm{Der}_F(K) = 0$.

Proof. Suppose that $D \in \mathrm{Der}_F(K)$. If $a \in K$, let $p(x) = \min(F, a)$, a separable polynomial over F. Then

$$0 = D(p(a)) = p'(a) D(a)$$

by Lemma 23.5. Since p is separable over F, the polynomials p and p' are relatively prime, so $p'(a) \neq 0$. Therefore, $D(a) = 0$, so D is the zero derivation. □

Corollary 23.7 *Let* $k \subseteq F \subseteq K$ *be fields, and suppose that* K/F *is a finite separable extension. Then each* k*-derivation on* F *extends uniquely to a* k*-derivation on* K.

Proof. The uniqueness is a consequence of Proposition 23.6. If D_1 and D_2 are k-derivations of K with the same restriction to F, then $D_1 - D_2 \in \mathrm{Der}_F(K)$, so $D_1 = D_2$. We now show that any derivation $D \in \mathrm{Der}_k(F)$ can be extended to a derivation D' on K. We can write $K = F(u)$ for some u separable over F. Let $p(x) = \min(F, u)$, and say $p(t) = \sum \beta_i t^i$. We first define $D'(u)$ by

$$D'(u) = -\frac{\sum_i D(\beta_i) u^i}{p'(u)}.$$

To define D' in general, if $v \in K$, say $v = f(u)$ for some $f(t) \in F[t]$. If $f(t) = \sum_i a_i t^i$, define D' on K by

$$D'(v) = f'(u)D'(u) + \sum_i D(a_i)u^i.$$

These formulas are forced upon us by the requirement that D' is an extension of D. The verification that D' is indeed a well-defined derivation on K is straightforward but tedious and will be left to the reader. □

The converse of this proposition is also true, which we will verify shortly. To do this, we must look at inseparable extensions.

Proposition 23.8 *Suppose that* $\operatorname{char}(F) = p > 0$, *and let* $K = F(a)$ *be purely inseparable over* F. *If* $K \neq F$, *then* $\operatorname{Der}_F(K)$ *is a one-dimensional* K*-vector space.*

Proof. Define $D : K \to K$ by $D(f(a)) = f'(a)$. We need to show that D is well defined. Let $p(x) = \min(F, a)$. Then $p(x) = x^{p^m} - \alpha$ for some $m \in \mathbb{N}$ and some $\alpha \in F$. If $f(a) = g(a)$, then p divides $f - g$, so $f(x) - g(x) = p(x)q(x)$ for some q. Taking derivatives, we have $f'(x) - g'(x) = p(x)q'(x)$, since $p'(x) = 0$. Therefore, $f'(a) = g'(a)$, so D is well defined. A short calculation shows that D is an F-derivation on K. If E is any derivation of K, then $E(f(a)) = f'(a)E(a)$ by Lemma 23.5, so E is a scalar multiple of D, namely $E = \beta D$ if $\beta = E(a)$. Therefore, $\operatorname{Der}_F(K)$ is spanned by D, so $\operatorname{Der}_F(K)$ is one dimensional as a K-vector space. □

We can now prove the converse of Proposition 23.6. This converse gives a test for separability by using derivations.

Corollary 23.9 *If* K *is an algebraic extension of* F *with* $\operatorname{Der}_F(K) = 0$, *then* K/F *is separable.*

Proof. Suppose that $\operatorname{Der}_F(K) = 0$, and let S be the separable closure of F in K. If $K \neq S$, then there is a proper subfield L of K containing S and an $a \in K$ with $K = L(a)$ and K/L purely inseparable. The previous proposition shows that $\operatorname{Der}_L(K) \neq 0$, so $\operatorname{Der}_F(K)$ is also nonzero, since it contains $\operatorname{Der}_L(K)$. This contradicts the assumption that $\operatorname{Der}_F(K) = 0$, so K is separable over F. □

We now consider transcendental extensions. First, we need a lemma that will allow us to work with polynomial rings instead of rational function fields.

Lemma 23.10 *Let* A *be an integral domain with quotient field* K. *Then any derivation on* A *has a unique extension to* K. *If* $D \in \operatorname{Der}_B(A)$ *for some subring* B *of* A, *then the unique extension of* D *to* K *lies in* $\operatorname{Der}_F(K)$, *where* F *is the quotient field of* B.

Proof. Let $D \in \mathrm{Der}(A)$. Define $D' : K \to K$ by

$$D'(a/b) = \frac{bD(a) - aD(b)}{b^2}$$

if $a, b \in A$ and $b \neq 0$. We first note that D' is well defined. If $a/b = c/d$, then $ad = bc$, so $aD(d) + dD(a) = bD(c) + cD(b)$. Thus, by multiplying both sides by bd and rearranging terms, we get

$$bd^2D(a) - bcdD(b) = b^2dD(c) - abdD(d).$$

Using the relation $ad = bc$, we can simplify this to

$$d^2\left(bD(a) - aD(b)\right) = b^2\left(dD(c) - cD(d)\right),$$

so

$$\frac{bD(a) - aD(b)}{b^2} = \frac{dD(c) - cD(d)}{d^2},$$

proving that D' is well defined. Checking that D' is a derivation is straightforward and will be left to the reader.

To verify uniqueness of extensions, suppose that D is a derivation on K. If $\alpha \in K$, we may write $\alpha = a/b$ with $a, b \in A$. Then

$$\begin{aligned} D(\alpha) &= D(ab^{-1}) \\ &= b^{-1}D(a) + aD(b^{-1}) \\ &= b^{-1}D(a) - ab^{-2}D(b), \end{aligned}$$

the final equality coming from Example 23.4. This formula shows that D is determined by its action on A. □

The following proposition determines the module of derivations for a purely transcendental extension of finite transcendence degree.

Proposition 23.11 *Suppose that $K = k(x_1, \ldots, x_n)$ is the rational function field over a field k in n variables. Then $\mathrm{Der}_k(K)$ is an n-dimensional K-vector space with basis $\{\partial/\partial x_i : 1 \leq i \leq n\}$.*

Proof. Let $f \in k[x_1, \ldots, x_n]$. If $D \in \mathrm{Der}_k(K)$, then by Lemma 23.5, we have $D(f) = \sum_i D(x_i)(\partial f/\partial x_i)$. Therefore, the n partial derivations $\partial/\partial x_i$ span $\mathrm{Der}_k(k[x_1, \ldots, x_n])$. Moreover, they are K-linearly independent; if $\sum_j a_j \partial/\partial x_j = 0$, then

$$0 = \sum_j a_j \frac{\partial x_i}{\partial x_j} = a_i.$$

This proves independence, so the $\partial/\partial x_i$ form a basis for $\mathrm{Der}_k(k[x_1, \ldots, x_n])$. Finally, a use of the quotient rule (Example 23.4) shows that the $\partial/\partial x_i$ form a basis for $\mathrm{Der}_k(K)$. □

We can generalize this theorem to any finitely generated, separable extension.

Theorem 23.12 *Suppose that K/k is a finitely generated, separable extension. Then* $\operatorname{trdeg}(K/k) = \dim_K(\operatorname{Der}_k(K))$. *If $\{x_1, \ldots, x_n\}$ is a separating transcendence basis for K/k and if $F = k(x_1, \ldots, x_n)$, then there is a basis $\{\delta_i : 1 \le i \le n\}$ for* $\operatorname{Der}_k(K)$ *with $\delta_i|_F = \partial/\partial x_i$ for each i.*

Proof. Let $\{x_1, \ldots, x_n\}$ be a separating transcendence basis for K/k, and set $F = k(x_1, \ldots, x_n)$. The extension K/F is finite and separable. By Corollary 23.7, for each i the derivation $\partial/\partial x_i$ extends uniquely to a derivation δ_i on K. We show that the δ_i form a basis for $\operatorname{Der}_k(K)$. It is easy to see that the δ_i are K-linearly independent, for if $\sum_i a_i\delta_i = 0$ with the $a_i \in K$, then

$$0 = \left(\sum_i a_i\delta_i\right)(x_j) = \sum_i a_j \frac{\partial x_j}{\partial x_i} = a_j$$

for each j. To show that the δ_i span $\operatorname{Der}_k(K)$, let D be a k-derivation of K, and let $a_i = D(x_i)$. Then $D - \sum_i a_i\delta_i$ is a derivation on K that is trivial on F. But $\operatorname{Der}_F(K) = 0$ by Proposition 23.6, so $D = \sum_i a_i\delta_i$. □

Differentials

Let $B \subseteq A$ be commutative rings. Then the *module of differentials* $\Omega_{A/B}$ is the A-module spanned by symbols da, one for each $a \in A$, subject to the relations

$$\begin{aligned} d\alpha &= 0, \\ d(ab) &= adb + bda \end{aligned}$$

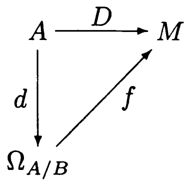
for $\alpha \in B$ and $a, b \in A$; that is, $\Omega_{A/B}$ is the A-module M/N, where M is the free A-module on the set of symbols $\{da : a \in A\}$ and N is the submodule generated by the elements

$$\begin{gathered} d\alpha, \\ d(a+b) - da - db, \\ d(ab) - (adb + bda) \end{gathered}$$

for $\alpha \in B$ and $a, b \in A$. The map $d : A \to \Omega_{A/B}$ given by $d(a) = da$ is a B-derivation on A by the definition of $\Omega_{A/B}$.

The module of differentials is determined by the following universal mapping property.

Proposition 23.13 *Suppose that $D : A \to M$ is a B-derivation from A to an A-module M. Then there is a unique A-module homomorphism $f : \Omega_{A/B} \to M$ with $f \circ d = D$; that is, $f(da) = D(a)$ for all $a \in A$. In other words, the following diagram commutes:*



Proof. Given D, we have an A-module homomorphism f defined on the free A-module on the set $\{da : a \in A\}$ into M that sends da to $D(a)$. Since D is a B-derivation, f is compatible with the defining relations for $\Omega_{A/B}$; hence, f factors through these relations to give an A-module homomorphism $f : \Omega_{A/B} \to M$ with $f(da) = D(a)$ for all $a \in A$. The uniqueness of f is clear from the requirement that $f(da) = D(a)$, since $\Omega_{A/B}$ is generated by $\{da : a \in A\}$. □

Corollary 23.14 *If $B \subseteq A$ are commutative rings and M is an A-module, then* $\mathrm{Der}_B(A, M) \cong \hom_A(\Omega_{A/B}, M)$.

Proof. This is really just a restatement of the universal mapping property for differentials. Define $\varphi : \mathrm{Der}_B(A, M) \to \hom_A(\Omega_{A/B}, M)$ by letting $\varphi(D)$ be the unique element f of $\hom_A(\Omega_{A/B}, M)$ that satisfies $f \circ d = D$. A short computation using the uniqueness part of the mapping property shows that φ is an A-module homomorphism. For injectivity, if $\varphi(D) = 0$, then the condition that $\varphi(D) \circ d = D$ shows that $D = 0$. Finally, for surjectivity, if $f \in \hom_A(\Omega_{A/B}, M)$, then setting $D = f \circ d$ yields $\varphi(D) = f$. □

If $M = A$, then the corollary shows that $\mathrm{Der}_B(A) \cong \hom_A(\Omega_{A/B}, A)$, the dual module to $\Omega_{A/B}$. The next corollary follows immediately from this observation.

Corollary 23.15 *If K is a field extension of F, then*

$$\dim_K(\Omega_{K/F}) = \dim_K(\mathrm{Der}_F(K)).$$

The following corollary is a consequence of the previous corollary together with Theorem 23.12.

Corollary 23.16 *If $\{x_1, \ldots, x_n\}$ is a separating transcendence basis for an extension K/k, then $\{dx_1, \ldots, dx_n\}$ is a K-basis for $\Omega_{K/k}$.*

Proof. Suppose that $\{x_1, \ldots, x_n\}$ is a separating transcendence basis for K/k. By Theorem 23.12, there is a basis $\{\delta_1, \ldots, \delta_n\}$ of $\mathrm{Der}_k(K)$ such that δ_i extends the derivation $\partial/\partial x_i$ on $k(x_1, \ldots, x_n)$. By the universal mapping property for differentials, there are $f_i \in \hom_K(\Omega_{K/k}, K)$ with $f_i(dx_j) = \delta_i(x_j)$ for each j. But, $\delta_i(x_j) = 0$ if $i \neq j$, and $\delta_i(dx_i) = 1$. Under the isomorphism $\mathrm{Der}_k(K) \cong \hom_K(\Omega_{K/k}, K)$, the δ_i are sent to the f_i, so

the f_i form a basis for $\hom_K(\Omega_{K/k}, K)$. The dual basis of $\Omega_{K/k}$ to the f_i is then $\{dx_1, \dots, dx_n\}$, so this set is a basis for $\Omega_{K/k}$. □

The converse of this corollary is also true, and the converse gives us a way to determine when a set of elements form a separating transcendence basis.

Proposition 23.17 *Suppose that K is a separably generated extension of k. If $x_1, \dots, x_n \in K$ such that $dx_1, \dots, dx_n$ is a K-basis for $\Omega_{K/k}$, then $\{x_1, \dots, x_n\}$ is a separating transcendence basis for K/k.*

Proof. Since K/k is separably generated, $n = \operatorname{trdeg}(K/k)$ by Theorem 23.12 and Corollary 23.15. Let $\{y_1, \dots, y_n\}$ be a separating transcendence basis for K/k. We will show that $\{x_1, \dots, x_n\}$ is also a separating transcendence basis by replacing, one at a time, a y_i by an x_j and showing that we still have a separating transcendence basis. The element x_1 is separable over $k(y_1, \dots, y_n)$, so there is an irreducible polynomial $p(t) \in k(y_1, \dots, y_n)[t]$ with $p(x_1) = 0$ and $p'(x_1) \neq 0$. We can write $p(t)$ in the form

$$p(t) = \frac{f_0}{g_0} + \frac{f_1}{g_1}t + \cdots + \frac{f_n}{g_n}t^n$$

with each $f_i, g_i \in k[y_1, \dots, y_n]$. By clearing denominators and dividing out the greatest common divisor of the new coefficients, we obtain a primitive irreducible polynomial $f(y_1, \dots, y_n, t)$ with $f(y_1, \dots, y_n, x_1) = 0$ and $(\partial f/\partial t)(y_1, \dots, y_n, x_1) \neq 0$. Let $P = (y_1, \dots, y_n, x_1)$. Taking differentials and using the chain rule yields

$$0 = \frac{\partial f}{\partial t}(P)dx_1 + \sum_{j=1}^{n} \frac{\partial f}{\partial y_j}(P)dy_j.$$

Consequently,

$$dx_1 = \sum_{j=1}^{n} -\frac{(\partial f/\partial y_j)(P)}{(\partial f/\partial t)(P)}dy_j.$$

The differential $dx_1 \neq 0$, so some $(\partial f/\partial y_j)(P) \neq 0$. By relabeling if necessary, we may assume that $(\partial f/\partial y_1)(P) \neq 0$. The equation $f(y_1, \dots, y_n, x_1) = 0$ shows that y_1 is algebraic over $k(x_1, y_2, \dots, y_n)$. Moreover, the condition $(\partial f/\partial y_1)(P) \neq 0$ implies that y_1 is separable over $k(x_1, y_2, \dots, y_n)$. Thus, each y_i is separable over $k(x_1, y_2, \dots, y_n)$, and since K is separable over $k(x_1, y_2, \dots, y_n)$, by transitivity the set $\{x_1, y_2, \dots, y_n\}$ is a separating transcendence basis for K/k.

Now, assume that for some $i \geq 1$, $\{x_1, \dots, x_i, y_{i+1}, \dots, y_n\}$ is a separating transcendence basis for K/k. Repeating the argument above for x_{i+1} in place of x_1, there is an irreducible primitive polynomial equation $g(Q) = 0$

with $(\partial g/\partial t_{n+1})(Q) \neq 0$, if $Q = (x_1, \ldots, x_i, y_{i+1}, \ldots, y_n, x_{i+1})$. This yields an equation

$$dx_{i+1} = \sum_{j=1}^{i} -\frac{(\partial g/\partial x_j)(Q)}{(\partial g/\partial t)(Q)} dx_j - \sum_{j=i+1}^{n} \frac{(\partial g/\partial y_j)(Q)}{(\partial g/\partial t)(Q)} dy_j.$$

The differentials $dx_1, \ldots, dx_n$ are K-independent, so some $(\partial g/\partial y_j)(Q) \neq 0$. Relabeling if necessary, we may assume that $(\partial g/\partial y_{i+1})(Q) \neq 0$. Consequently, y_{i+1} is separable over $k(x_1, \ldots, x_{i+1}, y_{i+2}, \ldots, y_n)$. As above, this means that $\{x_1, \ldots, x_{i+1}, y_{i+2}, \ldots, y_n\}$ is a separating transcendence basis for K/k. Continuing this procedure shows that $\{x_1, \ldots, x_n\}$ is a separating transcendence basis for K/k. □

Example 23.18 Let k_0 be a field of characteristic p, let $K = k_0(x, y)$ be the rational function field in two variables over k_0, and let $k = k_0(x^p, y^p)$. Then $\{x, y\}$ is algebraically dependent over k; in fact, K/k is algebraic. However, dx and dy are K-independent in $\Omega_{K/k}$; to see this, suppose that $adx + bdy = 0$ for some $a, b \in K$. The k_0-derivations $\partial/\partial x$ and $\partial/\partial y$ are actually k-derivations by the choice of k. By the universal mapping property for differentials, there are $f, g \in \hom_K(\Omega_{K/F}, K)$ with $f \circ d = \partial/\partial x$ and $g \circ d = \partial/\partial y$. Then $f(adx + bdy) = af(dx) + bf(dy) = a$ and $g(adx + bdy) = b$. Thus, $a = b = 0$, so dx and dy are K-independent. This shows that Proposition 23.17 is false if K/k is not separably generated.

The tangent space of a variety

Let $f(x, y)$ be a polynomial in $\mathbb{R}[x, y]$. The equation $f(x, y) = 0$ defines y implicitly as a function of x. If $P = (a, b)$ is a point on the curve $f = 0$, then, as long as the tangent line to the curve at P is not vertical, we have

$$\frac{dy}{dx}(a) = -\frac{\partial f/\partial x}{\partial f/\partial y}(P),$$

so the tangent line to the curve at P can be written in the form

$$\frac{\partial f}{\partial x}(P)(x - a) + \frac{\partial f}{\partial y}(P)(y - b) = 0.$$

This formula is valid even if the tangent line at P is vertical. To deal with vector subspaces, we define the *tangent space* to the curve $f = 0$ at P to be the set of solutions to the equation

$$\frac{\partial f}{\partial x}(P) \cdot x + \frac{\partial f}{\partial y}(P) \cdot y = 0.$$

This tangent space is a vector subspace of $\mathbb{R}^2$.

The curve $f = 0$ is nothing more than the set of $\mathbb{R}$-rational points of the $\mathbb{R}$-variety $Z(f)$. We can give a meaningful definition of the tangent space

to any k-variety, for any field k, by mimicking the case of real plane curves. Let V be a k-variety in C^n, where, as usual, C is an algebraically closed extension of k, and let $P \in V$. For $f \in k[x_1, \ldots, x_n]$, let

$$d_P f = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(P) x_i.$$

The linear polynomial $d_P f$ is called the *differential* of f at P.

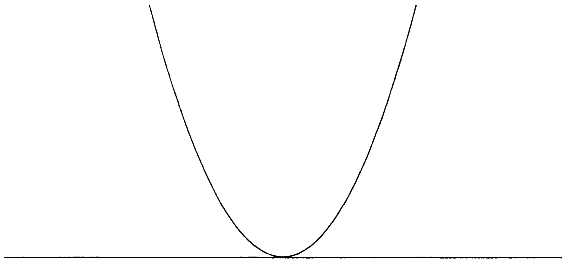
Definition 23.19 *If V is a k-variety, then the tangent space $T_P(V)$ to V at P is the zero set $Z(\{d_P f : f \in I(V)\})$.*

Example 23.20 By the Hilbert basis theorem, any ideal of $k[x_1, \ldots, x_n]$ can be generated by a finite number of polynomials. Suppose that $I(V)$ is generated by $\{f_1, \ldots, f_r\}$. Then we show that $T_P(V) = Z(\{d_P f_1, \ldots, d_P f_r\}$. If $h = \sum g_i f_i$, then by the product rule,

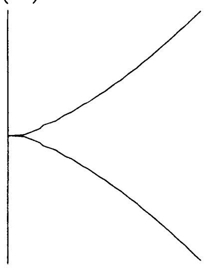
$$\begin{aligned} d_P h &= \sum g_i(P) d_P f_i + \sum d_P g_i \cdot f_i(P) \\ &= \sum d_i(P) d_P f_i. \end{aligned}$$

This shows that $d_P h$ is a linear combination of the $d_P f_i$ for any $h \in I(V)$.

Example 23.21 If $V = Z(y - x^2)$ and $P = (a, a^2)$, then $T_P(V) = Z(y + 2ax)$. If $P = (0,0)$ is the origin, then $T_P(V)$ is the x-axis.



Example 23.22 Let $V = Z(y^2 - x^3)$. If $P = (0,0)$, then $d_P f = 0$ for all $f \in I(V)$. Consequently, $T_P(V) = C^2$.



Example 23.23 Let $V = Z(x^2+y^2+z^2-1)$, and assume that $\text{char}(k) \neq 2$. If $P = (a, b, c)$ and $f = x^2 + y^2 + z^2 - 1$, then $d_P f = 2ax + 2by + 2cz$, so $T_P(V) = Z(ax + by + cz)$. Since $(a, b, c) \neq (0, 0, 0)$ for all $P \in V$, the tangent space $T_P(V)$ is a 2-dimensional vector space over C.

One of the uses of the tangent space is to define nonsingularity. To keep things as simple as possible, we first consider *hypersurfaces*; that is, varieties of the form $Z(f)$ for a single polynomial f.

Definition 23.24 *Let $V = Z(f)$ be a k-hypersurface. A point $P \in V$ is nonsingular, provided that at least one of the partial derivatives $\partial f/\partial x_i$ does not vanish at P; that is, P is nonsingular, provided that $d_P f \neq 0$. Otherwise, P is said to be singular. If every point on V is nonsingular, then V is said to be nonsingular.*

We can interpret this definition in other ways. The tangent space of $V = Z(f)$ at P is the zero set of $d_P f = \sum_i (\partial f/\partial x_i)(P)x_i$, so $T_P(V)$ is the zero set of a single linear polynomial. If $f \in k[x_1, \ldots, x_n]$, then $T_P(V)$ is either an $(n-1)$-dimensional vector space or is all of C^n, depending on whether $d_P f \neq 0$ or not. But, the point $P \in V$ is nonsingular if and only if $d_P f \neq 0$, so P is nonsingular if and only if $\dim_k(T_P(V)) = \dim(V) = n-1$, the latter equality from Example 22.8, and P is singular if $\dim_k(T_P(V)) > \dim(V)$.

Example 23.25 The parabola $Z(y - x^2)$ is a nonsingular curve, whereas $Z(y^2 - x^3)$ has a singularity at the origin. Every other point of $Z(y^2 - x^3)$ is nonsingular by an easy calculation. The sphere $Z(x^2 + y^2 + z^2 - 1)$ is also a nonsingular variety, provided that $\text{char}(k) \neq 2$.

For one application of the notion of nonsingularity, we point to Problem 6, which outlines a proof that the function field of the $\mathbb{C}$-variety $Z(y^2 - (x^3 - x))$ is not rational over $\mathbb{C}$.

We now look into nonsingularity for an arbitrary variety. Suppose that V is a k-variety, and let $f_1, \ldots, f_m$ be polynomials that generate the ideal $I(V)$. Let $P \in V$, and consider the Jacobian matrix

$$J(f_1, \ldots, f_m) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(P) & \cdots & \frac{\partial f_1}{\partial x_n}(P) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(P) & \cdots & \frac{\partial f_m}{\partial x_n}(P) \end{pmatrix}.$$

One interpretation of the definition of a nonsingular point on a hypersurface is that a point $P \in Z(f)$ is nonsingular if $\text{rank}(J(f)) = 1$, and P is singular if $\text{rank}(J(f)) = 0$. In other words, P is nonsingular if the rank of $J(f)$ is equal to $n - \dim(V)$.

Definition 23.26 *Suppose that V is an irreducible k-variety in C^n, and let $f_1, \ldots, f_m$ be generators of $I(V)$. If $P \in V$, then P is nonsingular if the rank of $J(f_1, \ldots, f_m)$ is equal to $n - \dim(V)$.*

The following proposition shows that $n - \dim(V)$ is an upper bound for the rank of the Jacobian matrix. Thus, a point is nonsingular, provided that the Jacobian matrix has maximal rank. We will call an irreducible k-variety V *absolutely irreducible* if the ideal $I(V)$ is an absolutely prime ideal of $k[x_1, \ldots, x_n]$.

Proposition 23.27 *Suppose that V is an absolutely irreducible k-variety in C^n. Let $P \in V$, and let $f_1, \ldots, f_m$ be generators of the ideal $I(V)$. Then* $\operatorname{rank}(J(f_1, \ldots, f_m)) \leq n - \dim(V)$.

Proof. We will prove this in a number of steps. Let K be the function field of V. The assumption that V is absolutely irreducible means that K/k is a regular extension, by Theorem 22.10. Therefore, K/k is separably generated, so $\operatorname{trdeg}(K/k) = \dim(\operatorname{Der}_k(K))$, and so $\dim(V) = \dim(\operatorname{Der}_k(K))$. The coordinate ring of V is $k[V] = k[x_1, \ldots, x_n]/I(V) = k[s_1, \ldots, s_n]$, where $s_i = x_i + I(V)$. Thus, $K = k(s_1, \ldots, s_n)$. Let $Q = (s_1, \ldots, s_n) \in K^n$. We first point out that

$$I(V) = \{f \in k[x_1, \ldots, x_n] : f(s_1, \ldots, s_n) = 0\}.$$

For $f \in I(V)$, let $d_Q f = \sum_{i=1}^n x_i(\partial f/\partial x_i)(Q)$. We view $d_Q f$ as a linear functional on K^n; that is, we view $d_Q f$ as a linear transformation from K^n to K defined by

$$(d_Q f)(\alpha_1, \ldots, \alpha_n) = \sum_{i=1}^n \alpha_i \frac{\partial f}{\partial x_i}(Q).$$

Let M be the subspace of $\hom_K(K^n, K)$ spanned by the $d_Q f$ as f ranges over $I(V)$. Now that we have given an interpretation of the differentials $d_Q f$ as linear functionals, we interpret derivations as elements of K^n. For $D \in \operatorname{Der}_k(K)$, we obtain an n-tuple $(D(s_1), \ldots, D(s_n))$. A k-derivation on K is determined by its action on the generators $s_1, \ldots, s_n$ of K/k. Therefore, the map $D \mapsto (D(s_1), \ldots, D(s_n))$ is a K-vector space injection from $\operatorname{Der}_k(K)$ to K^n. We denote by $\mathcal{D}$ the image of this transformation.

Next, we verify that an n-tuple $(\alpha_1, \ldots, \alpha_n)$ lies in $\mathcal{D}$ if and only if $d_Q f(\alpha_1, \ldots, \alpha_n) = 0$. One direction of this is easy. By the chain rule, we see that

$$\sum_{i=1}^n \frac{\partial f}{\partial x_i}(Q) D(s_i) = D(f(s_1, \ldots, s_n)) = 0$$

if $f \in I(V)$. For the other direction, suppose that $d_Q f(\alpha_1, \ldots, \alpha_n) = 0$. We define a derivation D on K with $D(s_i) = \alpha_i$ as follows. First, let D' be the derivation $D' : k[x_1, \ldots, x_n] \to K$ defined by $D' = \sum_i \alpha_i(\partial/\partial x_i)(Q)$; that is, $D'(f) = \sum_i \alpha_i(\partial f/\partial x_i)(Q)$. The condition on the α_i shows that $D'(f) = 0$ if $f \in I(V)$, so D' induces a k-derivation $D : k[V] \to K$ defined by $D(g + I(V)) = D'(g)$. The quotient rule for derivations shows that D

extends uniquely to a derivation on K, which we also call D. The definition of D' gives us $D(s_i) = \alpha_i$, so $(\alpha_1, \ldots, \alpha_n) \in \mathcal{D}$ as desired. Now that we have verified our claim, we use linear algebra. The subspace $\mathcal{D}$ of K^n is the set

$$\mathcal{D} = \{v \in K^n : d_Q f(v) = 0 \text{ for all } d_Q f \in M\}.$$

From linear algebra, this implies that $\dim(\mathcal{D}) + \dim(M) = n$. Since

$$\dim(M) = \dim(\mathrm{Der}_k(K)) = \dim(V),$$

we get $\dim(\mathcal{D}) = n - \dim(V)$.

The final step is to verify that $\dim(\mathcal{D}) = \mathrm{rank}(J')$, where J' is the matrix $((\partial f_i/\partial x_j)(Q))$, and that $\mathrm{rank}(J') \geq \mathrm{rank}(J)$, if J is the Jacobian matrix $((\partial f_i/\partial x_j)(P))$. This will show that

$$\mathrm{rank}(J) \leq \mathrm{rank}(J') = n - \dim(V),$$

our desired result. The first of these claims is easy. The space $\mathcal{D}$ is spanned by the $d_Q f_i$, since the f_i generate the ideal $I(V)$. The ith row of J' is the matrix representation of the linear transformation $d_Q f_i$, so the rank of J' is the dimension of the space spanned by the $d_Q f_i$; in other words, $\mathrm{rank}(J) = \dim(\mathcal{D})$. For the inequality $\mathrm{rank}(J') \geq \mathrm{rank}(J)$, let $P = (a_1, \ldots, a_n) \in V$. There is a homomorphism $\varphi : k[x_1, \ldots, x_n] \to C$ with $\varphi(x_i) = a_i$. Since $P \in V$, we have $f(P) = 0$ for all $f \in I(V)$, so $I(V) \subseteq \ker(\varphi)$. We get an induced map $\overline{\varphi} : k[V] \to C$ that sends s_i to a_i. Under this map $(\partial f_i/\partial x_j)(Q)$ is sent to $(\partial f_i/\partial x_j)(P)$. If $\mathrm{rank}(J') = r$, then the rows of J' are linear combinations of some r rows of J'. Viewing $\overline{\varphi}$ as a map on matrices, since $\overline{\varphi}(J') = J$ the rows of J are linear combinations of the corresponding r rows of J. Thus, the rank of J is at most r, so $\mathrm{rank}(J') \geq \mathrm{rank}(J)$. This finishes the proof. □

As a consequence of the proof of this proposition, we obtain a relation between the dimension of the tangent space $T_P(V)$ and of V.

Corollary 23.28 *Let V be an absolutely irreducible k-variety, and let $P \in V$. Then $\dim(T_P(V)) \geq \dim(V)$, and $\dim(T_P(V)) = \dim(V)$ if and only if P is nonsingular.*

Proof. The tangent space $T_P(V)$ is the set

$$T_P(V) = \{Q \in C^n : d_P f(Q) = 0 \text{ for all } f \in I(V)\}.$$

Using the notation of the proof of the previous proposition, the map $\overline{\varphi}$ induces a map on differentials that sends $d_Q f$ to $d_P f$. If $N = \{d_P f : f \in I(V)\}$, viewed as a subspace of $\hom_C(C^n, C)$, then by linear

algebra, we have $\dim(N) + \dim(T_P(V)) = n$. However, $\overline{\varphi}$ sends M to N, so $\dim(M) \geq \dim(N)$; hence,

$$\begin{aligned}\dim(T_P(V)) &= n - \dim(N) \geq n - \dim(M) \\ &= n - \dim(V).\end{aligned}$$

Moreover, $\dim(T_P(V)) = \operatorname{rank}(J)$ by the same argument that shows $\dim(\mathcal{D}) = \operatorname{rank}(J')$. Therefore, we get equality above exactly when $\operatorname{rank}(J') = \operatorname{rank}(J)$ or when $\operatorname{rank}(J) = n - \dim(V)$. However, this is true if and only if P is nonsingular, by the definition of nonsingularity. □

Let k be a field, and let C be an algebraically closed extension of k. In Example 22.4, we showed how one can obtain an irreducible k-variety from a finitely generated field extension of k. This map is not the inverse of the map that associates to each irreducible k-variety V the function field $k(V)$. In that example, we saw that the nonsingular curve $y = x^2$ has the same function field as the singular curve $y^2 = x^3$. However, nonsingularity is not the only problem. We have only talked about *affine* varieties; that is, varieties inside the affine space C^n. In algebraic geometry, one usually works with *projective* varieties. It is proved in many algebraic geometry books that there is a 1–1 correspondence between finitely generated regular extensions of k of transcendence degree 1 and nonsingular projective curves. Moreover, if we work over $\mathbb{C}$, then there is also a 1–1 correspondence between finitely generated extensions of $\mathbb{C}$ of transcendence degree 1 and Riemann surfaces. The interested reader can find the correspondence between nonsingular projective curves and extensions of transcendence degree 1 in Section 1.6 of Hartshorne [11] and can find the connection with Riemann surfaces in Chevalley [4].

Problems

1. Let K be a separable extension of F that is not necessarily algebraic. Show that any derivation on F extends to a derivation on K.

2. If K is a finite separable extension of F, show that there is a K-vector space isomorphism $\operatorname{Der}_k(F) \otimes_F K \cong \operatorname{Der}_k(K)$.

3. Let K be a finite purely inseparable extension of F with $\operatorname{char}(F) = p$ and $F^p \subseteq K$. Recall that a p-basis of K/F is a set $\{a_1, \dots, a_n\}$ of elements of K with

$$F \subset F(a_1) \subset \cdots \subset F(a_n) = K,$$

and the p-dimension of K/F is n. Show that $\dim(\operatorname{Der}_F(K))$ is equal to the p-dimension of K/F.
(See Problem 12 of Section 4 and Problem 19 of Appendix D for more on p-dependence.)

4. Let V be a C-variety in C^n, where C is algebraically closed.

   (a) If $P \in V$, show that each $d_P f$ for $f \in C[x_1, \ldots, x_n]$ defines a linear transformation from C^n to C, so $d_P f$ restricts to a linear transformation from $T_P(V)$ to C.

   (b) Let $M_P = \{f \in C[x_1, \ldots, x_n] : f(P) = 0\}$. Show that the function $d_P : M_P \to \hom_C(T_P(V), C)$ is a C-vector space homomorphism with kernel M_P^2. Show that d_P is surjective, and conclude that M_P/M_P^2 is isomorphic to $\hom_C(T_P(V), C)$.

   (c) Show that $T_P(V)$ is isomorphic to $\hom_C(M_P/M_P^2, C)$.

5. Let V and W be k-varieties, and suppose that $\varphi : V \to W$ is a morphism. Show that φ induces a homomorphism $T_P(V) \to T_{\varphi(P)}(W)$.

6. Let $X \subseteq \mathbb{C}^2$ be the zero set of $y^2 - x^3 + x$. In this problem, we will show that the function field $\mathbb{C}(Y)$ is not rational over $\mathbb{C}$. In order to do this, we need the following result: If Y is an irreducible nonsingular curve in $\mathbb{C}^2$ such that $\mathbb{C}(Y)/\mathbb{C}$ is rational, then $\mathbb{C}[Y]$ is a unique factorization domain. Verify that $\mathbb{C}(X)$ is not rational over $\mathbb{C}$ by verifying the following steps.

   (a) Show that X is an irreducible nonsingular curve.

   (b) Let $F = \mathbb{C}(x) \subseteq K$. Show that K/F is a degree 2 extension. If σ is the nonidentity F-automorphism of K, show that $\sigma(y) = -y$. Conclude that $\sigma(A) \subseteq A$, where $A = \mathbb{C}[X]$.

   (c) Let $N = N_{K/F}$ be the norm map from K to F. Show that $N(a) \in k[x]$ for all $a \in A$.

   (d) Using the norm map, show that the units in A are merely the nonzero elements of $\mathbb{C}$. Write x and y for the images of x and y in A, and show that x and y are irreducible elements of A, and conclude that A is not a unique factorization domain. From this, conclude that $\mathbb{C}(X)$ is not rational over $\mathbb{C}$.

   (Note: To prove that $\mathbb{C}[Y]$ is a unique factorization domain for a rational curve Y requires more geometry than we have developed here. Problem 6.1 of Chapter I in Hartshorne [11] outlines a proof of this fact along with the steps above. For an alternative proof that does not require geometry but does use valuation theory, see §1 of [5].)

Appendix A
Ring Theory

The following appendices present some of the background material used in this book. In this appendix, we present the aspects of ring theory that we need in this book. We go into detail about unique factorization domains and polynomials in multiple variables, mostly for Chapter V. The irreducibility tests are used in the text mainly for dealing with polynomials over $\mathbb{Q}$ and over the rational function field $k(x)$ over a field k.

Throughout this book, we make the assumption that all rings have a multiplicative identity. We start off with a review of the characteristic of a ring. Let R be a ring. The characteristic of R, denoted $\text{char}(R)$, is the order of 1 in the additive group $(R, +)$, provided that this order is finite; if it is infinite, we set $\text{char}(R) = 0$. Here is an alternative description of $\text{char}(R)$. There is a map $\varphi : \mathbb{Z} \to R$ given by $\varphi(n) = n \cdot 1$, the sum of 1 with itself n times. It is clear that this map is a ring homomorphism. The kernel of φ is generated by a positive integer m, and m is precisely $\text{char}(R)$. Thus, $\mathbb{Z}/m\mathbb{Z}$ is isomorphic to a subring of R. Moreover, this subring is easily seen to be the unique minimal subring of R; recall that we assume that our rings have an identity. This ring is called the *prime subring* of R.

We remind the reader that all references to Theorems, Lemmas, etc., made in each appendix refer to that appendix unless it is explicitly stated that they come from a section of the main text.

1 Prime and Maximal Ideals

Let R be a commutative ring. A *prime ideal* of R is an ideal $P \neq R$, such that if $a, b \in R$ with $ab \in P$, then either $a \in P$ or $b \in P$. For example, if p is a prime number, then the ideal $p\mathbb{Z}$ is a prime ideal of $\mathbb{Z}$. A *maximal ideal* of R is an ideal $M \neq R$, such that if I is any ideal of R with $M \subseteq I \subseteq R$, then either $I = M$ or $I = R$; that is, M is maximal if M is not contained in any proper ideal other than itself. Again, if p is a prime number, then $p\mathbb{Z}$ is a maximal ideal of $\mathbb{Z}$. This can be seen from the fact that the gcd of two integers can be written as a linear combination of the integers. If $p\mathbb{Z} \subseteq I$ and $I \neq p\mathbb{Z}$, let $a \in I$ with $a \notin p\mathbb{Z}$. Then p does not divide a, so $\gcd(a, p) = 1$. Therefore, $1 = ax + py$ for some $x, y \in \mathbb{Z}$. This means that $1 \in I$, since $a, p \in I$; hence, $I = \mathbb{Z}$. This proves that $p\mathbb{Z}$ is indeed a maximal ideal of $\mathbb{Z}$.

Prime and maximal ideals can be characterized in terms of quotient rings. This characterization is often a very useful way to deal with these ideals.

Proposition 1.1 *Let R be a commutative ring with 1.*

1. *If P is a proper ideal of R, then P is a prime ideal of R if and only if R/P is an integral domain.*

2. *If M is a proper ideal of R, then M is a maximal ideal of R if and only if R/M is a field.*

Proof. Let P be a prime ideal. To show that R/P is an integral domain, suppose that $\alpha, \beta \in R/P$ with $\alpha\beta = 0$. Then $\alpha = a + P$ and $\beta = b + P$ for some $a, b \in R$. The condition $\alpha\beta = 0$ in R/P means $(a+P)(b+P) = 0+P$, so $ab \in P$. Since P is a prime ideal, either $a \in P$ or $b \in P$, so $a+P = 0+P$ or $b + P = 0 + P$. Thus, R/P is an integral domain. The converse follows from the same arguments; if R/P is an integral domain and $ab \in P$, then $(a+P)(b+P) = 0$ in R/P, so $a+P = 0+P$ or $b+P = 0+P$; thus, $a \in P$ or $b \in P$.

For the second statement, suppose that M is a maximal ideal of R. We need to show that each nonzero element of R/M is invertible. Take $a + M \in R/M$ with $a + M \neq 0 + M$. Then $a \notin M$, so the ideal $M + aR$ is properly larger than M. By maximality, this forces $M + aR = R$, so $1 = m + ar$ for some $m \in M$ and $r \in R$. Then $(a + M)(r + M) = 1 + M$, since $m \in M$. Therefore, $a + M$ is invertible, so R/M is a field. Conversely, suppose that R/M is a field. Let I be an ideal of R with $M \subset I \subseteq R$. We need to show that $I = R$. Let $a \in I - M$. Then $a + M \neq 0 + M$; hence, $a+M$ is invertible. Thus, there is a $b \in R$ with $(a+M)(b+M) = 1+M$, so $ab - 1 \in M$. Since $a \in I$ and $M \subseteq I$, this forces $1 \in I$, so $I = R$. Therefore, M is a maximal ideal of R. □

From this proposition, we see that any maximal ideal is prime, but the converse may not be true. Our main use of these concepts will be for the

study of polynomials. It follows from the results of Section 3 below that if F is a field and $R = F[x]$ is the ring of polynomials over F, then any prime ideal of R is maximal and is generated by an irreducible polynomial.

Example 1.2 By calculations similar to those before the proposition, one can show that an ideal $a\mathbb{Z}$ of $\mathbb{Z}$ is a prime ideal if and only if a is a prime number. Moreover, an ideal I of $\mathbb{Z}$ is maximal if and only if I is prime.

Let $R = \mathbb{Z}[x]$, the ring of polynomials in x over $\mathbb{Z}$. The ideal xR is prime, since $R/xR \cong \mathbb{Z}$ is an integral domain. Moreover, xR is not maximal, since R/xR is not a field. Equivalently, xR is not maximal, since xR is properly contained in the proper ideal $xR + 2R$ generated by x and 2.

The proposition above also gives us some information about the characteristic of a ring. If R is an integral domain, then the map $\varphi : \mathbb{Z} \to R$ that sends n to $n \cdot 1$ is a ring homomorphism, and $\text{im}(\varphi)$ is a subring of R. Thus, $\mathbb{Z}/\ker(\varphi)$ is an integral domain, so $\ker(\varphi)$ is a prime ideal. But $\ker(\varphi)$ is generated by $\text{char}(R)$, so $\text{char}(R)$ is either 0 or a prime number.

2 Unique Factorization Domains

The main ring theoretic properties about polynomials we require in Galois theory are that the ring $F[x]$ of polynomials in a variable x over a field F be a principal ideal domain (PID) and be a unique factorization domain (UFD). While these facts can be proved relatively easily, we go into some detail about UFDs primarily to deal with polynomials in more than one variable, a case we need in Chapter V.

Let R be an integral domain. If $a, b \in R$, we say that a *divides* b if $b = ac$ for some $c \in R$. A nonunit $a \in R$ is said to be *irreducible* if whenever $a = bc$, then either b or c is a unit. A nonunit $a \in R$ is said to be *prime* if whenever a divides bc, then a divides b or a divides c. Equivalently, a is prime if the principal ideal aR is a prime ideal. If a is prime, then we show that a is irreducible. If $a = bc$, then a divides bc; hence, a divides b or c. If a divides b, then $b = ad$ for some d. Consequently, $1 = dc$, so c is a unit. On the other hand, if a divides c, then the same argument shows that b is a unit. However, irreducible elements need not be prime. Perhaps the easiest example is in the ring $\mathbb{Z}[\sqrt{-5}]$. In this ring, $6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$. With some calculation, we can see that 2 is irreducible and that 2 does not divide either of $1 \pm \sqrt{-5}$. Since 2 divides their product, 2 is not prime in $\mathbb{Z}[\sqrt{-5}]$.

Definition 2.1 *An integral domain R is a unique factorization domain (UFD) if every nonzero nonunit of R can be factored uniquely into a product of irreducible elements.*

Some words about this definition are in order. What does it mean to factor an element uniquely? In $\mathbb{Z}$, the integer 6 factors as $6 = 2 \cdot 3$ and as $6 = (-2) \cdot (-3)$. The four elements ± 2 and ± 3 are prime according to our definition. This means we have to be more precise in our meaning. If $a, b \in R$ such that a divides b and b divides a, we say that a and b are *associates*. Equivalently, a and b are associates if $aR = bR$ (see Problem 4). Therefore, two associates differ by multiplication by a unit. In studying divisibility, units are trivial; hence, we would like not to have to worry about them. Therefore, we say that an element a factors uniquely into a product of irreducible elements if a is a product of irreducibles, and if $a = \pi_1^{e_1} \cdots \pi_n^{e_n} = \theta_1^{f_1} \cdots \theta_m^{f_m}$ with each π_i and θ_j irreducible, then $n = m$, and after reordering, $e_i = f_i$ and $\pi_i R = \theta_i R$ for each i. Therefore, unique factorization means unique up to multiplication by units.

While irreducible elements may not be prime, they are in a UFD. This fact will be used frequently when dealing with polynomial rings.

Lemma 2.2 *Let R be a UFD. If $\pi \in R$ is irreducible, then π is prime.*

Proof. Suppose that π divides ab. Then $ab = \pi c$ for some c. If $c = \theta_1^{f_1} \cdots \theta_m^{f_m}$ is the factorization of c into irreducibles, then $\pi c = \pi \theta_1^{f_1} \cdots \theta_m^{f_m}$ is the factorization of $\pi c = ab$ into irreducibles. However, if we look at the factorization of a and b, by uniqueness π must occur in one of these factorizations. Therefore, π divides a or π divides b, so π is prime. □

There are some equivalent definitions of a UFD. Some of these are addressed in the problems at the end of this appendix. One characterization, due to Kaplansky, is presented now, and we will use it to show that the ring of polynomials over a field is a UFD. This is a prime ideal theoretic characterization of a UFD, and is quite useful in proving facts about UFDs. Another characterization is that a ring is a UFD if and only if each nonunit can be factored into a product of primes. We use this characterization in the proof of the following theorem, although we leave its proof to the reader (Problem 9).

Theorem 2.3 (Kaplansky) *Let R be an integral domain. Then R is a UFD if and only if each nonzero prime ideal of R contains a nonzero principal prime ideal.*

Proof. Suppose that R is a UFD, and let P be a nonzero prime ideal of R. If $a \in P$ with $a \neq 0$, let $a = \pi_1 \cdots \pi_n$ be a prime factorization of a. Since P is a prime ideal and $a \in P$, one of the π_i must be in P. Therefore, P contains the principal prime ideal (π_i).

Conversely, suppose that every nonzero prime ideal of R contains a nonzero principal prime ideal. Let

$$\mathcal{S} = \{a \in R - \{0\} : a \text{ is a unit or } a \text{ factors into a product of primes}\}.$$

If $\mathcal{S} = R - \{0\}$, then R is a UFD by Problem 9. If not, there is an $a \in R - \mathcal{S}$ with $a \neq 0$. Let I be an ideal of R containing a that is maximal under inclusion among the ideals disjoint from $\mathcal{S}$. Such an ideal exists by an easy application of Zorn's lemma. We claim that I is a prime ideal. Assuming this for the moment, by hypothesis I contains a prime element π. However, $\pi \in \mathcal{S}$, since π is prime. But $I \cap \mathcal{S} = \varnothing$, a contradiction. Therefore, $\mathcal{S} = R - \{0\}$, and so R is a UFD.

We are then left with showing that I is prime. First, we note that $\mathcal{S}$ is closed under multiplication. If I is not prime, then there are $b, c \in R - I$ with $bc \in I$. Then $I + bR$ and $I + cR$ are larger than I, so, by maximality, both intersect $\mathcal{S}$. Say $x \in \mathcal{S} \cap (I + bR)$ and $y \in \mathcal{S} \cap (I + cR)$. If $x = u_1 + br_1$ and $y = u_2 + cr_2$ with $u_i \in I$ and $r_i \in R$, then $xy = u_1(u_2 + cr_2) + bcr_1r_2 \in I$, since $bc \in I$. But $xy \in \mathcal{S}$, since $\mathcal{S}$ is closed under multiplication. This forces $\mathcal{S} \cap I \neq \varnothing$, a contradiction. Therefore, I is prime. □

We finish this section with a short discussion of greatest common divisors.

Definition 2.4 *Let R be a UFD. If $a, b \in R$ are nonzero elements, then a greatest common divisor of a and b is an element d such that*

1. *d divides a and d divides b;*
2. *if e divides a and e divides b, then e divides d.*

The gcd of two elements is not unique if it exists. However, by the second condition in the definition, it follows that any two gcds of a and b are associates. We often abuse language and call an element d the gcd of a and b, and write $d = \gcd(a, b)$.

The definition of gcd makes perfect sense in any commutative ring. The difficulty is that a gcd of two elements need not exist, as shown in Problem 11. However, if R is a UFD, then we can see that a gcd always exists. In fact, if $b = \pi_1^{e_1} \cdots \pi_n^{e_n}$ and $a = \pi_1^{f_1} \cdots \pi_n^{f_n}$ is the factorization of a and b into irreducibles, where e_i, f_i can be 0, and if $g_i = \min\{e_i, f_i\}$, then a gcd of a and b is $\pi_1^{g_1} \cdots \pi_n^{g_n}$. If 1 is a gcd of a and b, we say that a and b are *relatively prime*. Unlike in the integers, in a UFD a gcd of two elements need not be a linear combination of the elements. An example of this appears in Problem 15. However, if R is a PID and $d = \gcd(a, b)$, then $dR = aR + bR$, so d is a linear combination of a and b (Problem 17).

The definition of gcd can be extended to any finite set of elements instead of just two elements. An element $d \in R$ is a gcd of $a_1, \ldots, a_n$ if d divides each a_i, and any e that divides all a_i also divides d. In a UFD, the gcd of any finite set of elements does exist. Moreover, a gcd can be calculated by recursion from the equation $\gcd(a_1, \ldots, a_n) = \gcd(a_1, \gcd(a_2, \ldots, a_n))$ (see Problem 16).

3 Polynomials over a Field

Let R be a ring. We denote by $R[x]$ the ring of polynomials in one variable over R. Given a polynomial $f(x) = \sum_{i=0}^{n} r_i x^i$ with $r_n \neq 0$, the *degree* of f is defined to be n. For convenience, we define the degree of the zero polynomial to be $-\infty$. While our primary interest is in polynomials over a field, we state a number of results for polynomials over an integral domain.

Lemma 3.1 *Let R be an integral domain. If $f(x), g(x) \in R[x]$ are nonzero, then* $\deg(fg)) = \deg(f) + \deg(g)$. *Consequently, $R[x]$ is an integral domain.*

Proof. Let $f(x) = \sum_{i=0}^{n} a_i x^i$ and $g(x) = \sum_{i=0}^{m} b_i x^i$ with $\deg(f) = n$ and $\deg(g) = m$. Therefore, $a_n, b_m \neq 0$. The product of f and g is $f(x)g(x) = \sum_{i=0}^{n+m} (\sum_{j+k=i} a_j b_k) x^i$. Clearly, all coefficients past degree $n + m$ are 0. The coefficient of x^{n+m} is $a_n b_m$, which is nonzero, since both a_n and b_m are nonzero. This proves the degree formula. Moreover, it shows that fg cannot be the zero polynomial unless $f = 0$ or $g = 0$; hence, $R[x]$ is an integral domain. □

If either $f = 0$ or $g = 0$, then the degree formula still holds with the convention that $\deg(0) = -\infty$, given that addition is defined by $n + (-\infty) = -\infty$ for all integers n.

The main theorem for polynomials over a field is the division algorithm. Since this result holds in more generality and is useful in its full version, we give the full version here.

Theorem 3.2 (Division Algorithm) *Let R be an integral domain. Let $f(x), g(x) \in R[x]$ with $g(x) \neq 0$, and suppose that the leading coefficient of g is a unit in R. Then there are unique polynomials $q(x), r(x) \in R[x]$ satisfying $f(x) = q(x)g(x) + r(x)$ and* $\deg(r(x)) < \deg(g(x))$.

Proof. This argument is almost the same as the proof of the division algorithm for polynomials with real number coefficients. We first show the existence of q and r with the desired properties, then we prove the uniqueness. Let

$$\mathcal{S} = \{f(x) - q(x)g(x) : q(x) \in R[x]\}.$$

The set $\mathcal{S}$ is clearly nonempty. Let $r(x) \in \mathcal{S}$ be a polynomial of minimal degree in $\mathcal{S}$. Then $f = qg + r$. If $\deg(r) \geq \deg(g)$, then say $r(x) = \sum_{i=0}^{n} a_i x^i$ and $g(x) = \sum_{i=0}^{m} b_i x^i$ with $a_n, b_m \neq 0$ and $n \geq m$. If $q_1(x) = q(x) - a_n b_m^{-1} x^{n-m}$, which makes sense since b_m is assumed to be a unit in R, then

$$f(x) - q_1(x)g(x) = r(x) - a_n b_m^{-1} x^{n-m} g(x),$$

which has degree less than n, since the coefficient of x^n is 0. Consequently, this polynomial is in $\mathcal{S}$ and has smaller degree than $r(x)$. This is a contradiction, which forces $n < m$.

For uniqueness, suppose that there are $q(x), q_1(x)$ and $r(x), r_1(x) \in R[x]$ with $f = qg + r$ and $f = q_1 g + r_1$, and with $\deg(r), \deg(r_1) < \deg(g)$. Then $g\,(q_1 - q) = r - r_1$. If $q_1 \neq q$, then the degree of $g(q_1 - q)$ is at least $\deg(g)$, which is larger than $\deg(r - r_1)$. This contradiction shows that $q_1 = q$, which forces $r = r_1$. This proves the uniqueness. □

We state the usual division algorithm separately for emphasis.

Corollary 3.3 *If F is a field and if $f(x), g(x) \in F[x]$ with $g(x) \neq 0$, then there are unique polynomials $q(x), r(x) \in F[x]$ satisfying $f(x) = q(x)g(x) + r(x)$ and $\deg(r(x)) < \deg(g(x))$.*

The division algorithm yields the fact that $F[x]$ is a PID. From this, we will see that $F[x]$ is a UFD.

Corollary 3.4 *If F is a field, then $F[x]$ is a PID.*

Proof. Let I be an ideal of $F[x]$. If $I = \{0\}$, then I is generated by 0. If $I \neq 0$, take $g(x) \in I - \{0\}$ of minimal degree. If $f(x) \in I$, by the division algorithm there are polynomials q and r with $f = qg + r$ and $\deg(r) < \deg(g)$. Since I is an ideal, $r = f - qg \in I$. Minimality of $\deg(g)$ forces $r = 0$, which shows that f is in the ideal generated by g. Therefore, $I = (g)$ is principal, so $F[x]$ is a PID. □

We can now use Kaplansky's theorem to give an easy proof that $F[x]$ is a UFD.

Lemma 3.5 *If R is a PID, then R is a UFD. In particular, if F is a field, then $F[x]$ is a PID.*

Proof. Suppose that R is a PID. If P is a prime ideal of R, then P is principal, say $P = (\pi)$. Therefore, P contains the principal prime ideal (π). By Theorem 2.3, R is a UFD. □

The following fact will be used early in Chapter I.

Corollary 3.6 *Let F be a field. If $p(x) \in F[x]$, then the principal ideal $(p(x))$ is a maximal ideal of $F[x]$ if and only if $p(x)$ is irreducible. Consequently, any prime ideal of $F[x]$ is maximal.*

Proof. This really is a fact about PIDs, as the proof will show. Suppose that $p(x)$ is irreducible. Let M be a maximal ideal of $F[x]$ containing $p(x)$. Since $F[x]$ is a PID, $M = (f(x))$ for some polynomial $f(x)$. Then f divides p since $(p) \subseteq (f)$. But p is irreducible, so p has no divisors other than units or associates. Since $(f) = M \neq F[x]$, we see that f is not a unit; hence, f is an associate to p. Therefore, $(f) = (p)$, so (p) is maximal. Conversely,

suppose that (p) is maximal. If p is not irreducible, then $p = fg$ with both f and g nonconstant polynomials. Then $(p) \subset (f) \subset R$. This contradicts maximality, so p is irreducible. If M is a prime ideal, then $M = (p)$ for some irreducible polynomial by arguments similar to those just given; hence, M is maximal. □

4 Factorization in Polynomial Rings

The goal of this section is to show that $R[x]$ is a UFD whenever R is a UFD. However, we have some work to do in order to prove this.

Definition 4.1 *Let R be a UFD, and let $f(x) \in R[x]$. The content $c(f)$ of f is the gcd of the coefficients of f. If the content of f is 1, then f is said to be primitive.*

The following lemma is easy to prove, but we will use it in a number of places in this book. The proof follows immediately from the definition of addition and multiplication in polynomial rings and quotient rings; this will be left to the reader.

Lemma 4.2 *Let R be a ring, and let I be an ideal of R. Then the map $\varphi : R[x] \to R/I[x]$ given by $\varphi(\sum_i a_i x^i) = \sum_i (a_i + I)x^i$ is a surjective ring homomorphism.*

In particular, if p is a prime number and $\overline{a}$ represents the equivalence class of a modulo p, then the map $\mathbb{Z}[x] \to \mathbb{F}_p[x]$ given by $\sum_i a_i x^i \to \sum_i \overline{a_i} x^i$ is a ring homomorphism.

Proposition 4.3 *Let R be a UFD, and let $f, g \in R[x]$. Then $c(fg) = c(f)c(g)$. In particular, if f and g are primitive, then fg is primitive.*

Proof. We may write $f(x) = c(f)f_1(x)$ and $g(x) = c(g)g_1(x)$ for some primitive polynomials f_1 and g_1. So, $fg = c(f)c(g) \cdot f_1 g_1$. If we prove that the product of primitive polynomials is primitive, then we will have proved the proposition. So, suppose that f and g are primitive. If fg is not primitive, then there is a prime element π that divides all of the coefficients of fg. Consider the polynomial ring $R/(\pi)[x]$ over $R/(\pi)$. Since π is a prime element, $R/(\pi)$ is an integral domain. Let $\overline{f}$ and $\overline{g}$ be the images of f and g in $R/(\pi)[x]$. Since f and g are primitive, π does not divide all the coefficients of f or g, so $\overline{f} \neq 0$ and $\overline{g} \neq 0$. Therefore, $\overline{f} \cdot \overline{g} = \overline{fg}$ by Lemma 4.2, and $\overline{f} \cdot \overline{g} \neq 0$, since $R/(\pi)[x]$ is an integral domain. However, if π divides all the coefficients of fg, then $\overline{fg} = 0$, a contradiction. Therefore, fg is indeed primitive. □

The following theorem is perhaps the most important result about polynomials over a UFD.

Theorem 4.4 (Gauss' Lemma) *Let R be a UFD, let F be its quotient field, and let $f(x) \in R[x]$. Then f is irreducible over R if and only if f is primitive and irreducible over F.*

Proof. Suppose that f is primitive and irreducible over F. If f factors in $R[x]$ as $f = gh$ with neither g nor h a unit in $R[x]$, then since f is irreducible over F, either g or h must be a constant. But if g is a constant, g would divide all the coefficients of $gh = f$. This is impossible, since f is primitive, so f is irreducible over R.

Conversely, suppose that f is irreducible over R. Since we can write $f = c(f) \cdot f_1$ with f_1 primitive, $c(f)$ is a unit in $R[x]$; hence, $c(f)$ is a unit in R. Therefore, f is primitive. If f is not irreducible over F, then we can write $f = gh$ with $g, h \in F[x]$ both of degree at least 1. By using common denominators, we can write $gh = a/b \cdot g_1 h_1$, where $g_1, h_1 \in R[x]$ are primitive and $a, b \in R$ are relatively prime. Then $bf(x) = ag_1(x)h_1(x)$. By Proposition 4.3, we have $b = c(bf) = a$, since g_1 and h_1 are primitive. But this contradicts $\gcd(a, b) = 1$ unless a and b are both units in R. If a and b are units in R, then $f = (ag_1)(1/b \cdot h_1)$ is a nontrivial factorization of f in $R[x]$, which contradicts the assumption that f is irreducible over R. Therefore, f is indeed irreducible over F. □

We can now prove that $R[x]$ is a UFD if R is.

Theorem 4.5 *If R is a UFD, then $R[x]$ is a UFD.*

Proof. We give here a somewhat nonstandard proof of this theorem, making use of Theorem 2.3. This proof is easier to understand if the reader has some experience in localization. Some of the details in this proof are left to Problem 18. A more standard proof of this fact can be found in Hungerford [13, Thm. 3.11.1] or Herstein [12, Thm. 3.11.1].

Let Q be a nonzero prime ideal of $R[x]$. By Theorem 2.3, we wish to show that Q contains a nonzero prime element. If $P = Q \cap R$, then P is a prime ideal in R. If $P \neq 0$, then P contains a nonzero prime element π of R, which is also a prime element of $R[x]$. Therefore, Q contains a nonzero prime element of $R[x]$. The more difficult case is if $P = 0$, which we now consider. Let F be the quotient field of R. Then $F[x]$ is a UFD, as we have already seen. Let $Q' = QF[x]$, the ideal of $F[x]$ generated by Q. Then Q' is a prime ideal, since $P = 0$ (see Problem 18). Because $F[x]$ is a UFD, there is a polynomial $f(x) \in Q'$ that is irreducible in $F[x]$. We can write $f(x) = \frac{a}{b}g(x)$ with $g(x) \in R[x]$ a primitive polynomial and $a, b \in R$. Since a/b is a unit in F, we have $g(x) \in Q'$. Furthermore, $g(x)$ is irreducible in $F[x]$, since $g(x)$ is an associate to $f(x)$. Therefore, $g(x)$ is irreducible over R since $g(x)$ is primitive, by Gauss' lemma. But $g(x) \in Q' \cap R[x]$, so $g(x) \in Q$ (see Problem 18 again). Finally, we need to show that $g(x)$ is prime in $R[x]$, which will finish the proof. We see that $g(x)F[x] \cap R[x] = g(x)R[x]$, since $g(x) \in R[x]$ (see Problem 18). However, the ideal $g(x)F[x]$ is prime, since

$F[x]$ is a UFD and $g(x)$ is irreducible. Thus, $g(x) \in R[x]$ is prime, since the intersection of a prime ideal of $F[x]$ with $R[x]$ is a prime ideal of $R[x]$. □

Corollary 4.6 *If F is a field, then $F[x_1, \ldots, x_n]$ is a UFD for any n.*

Proof. This follows by induction on n and the previous theorem, the $n = 1$ case having been proven earlier. □

More generally, if F is a field and X is any set of variables, possibly infinite, then the polynomial ring $F[X]$ is a UFD. A proof of this fact can be obtained from the following two points. First, any element of $F[X]$ is a polynomial in finitely many of the variables, and unique factorization holds for polynomials in finitely many variables, and second, adding more variables does not affect whether an element is irreducible.

5 Irreducibility Tests

It is hard in general to determine if a polynomial is irreducible over a field F. However, if F is the quotient field of a UFD, there are some simple tests that can determine when a polynomial is irreducible over F. While these tests may seem somewhat specialized, nonetheless they can be quite useful.

The first test is actually a test for roots, but it is also an irreducibility test for polynomials of degree 2 and 3. Let R be a UFD, and let F be its quotient field. Suppose that $f(x) = a_0 + \cdots + a_n x^n \in R[x]$. If $\alpha/\beta \in F$ is a root of $f(x)$ with $\gcd(\alpha, \beta) = 1$, then by multiplying the equation $f(\alpha/\beta) = 0$ by β^n, we obtain the equation

$$a_0\beta^n + a_1\alpha\beta^{n-1} + \cdots + a_{n-1}\alpha^{n-1}\beta + a_n\alpha^n = 0.$$

Therefore, $a_0\beta^n = -\alpha\left(a_1\beta^{n-1} + \cdots + a_{n-1}\alpha^{n-2}\beta + a_n\alpha^{n-1}\right)$. Since α is relatively prime to β, it follows that α divides a_0. By a similar manipulation, we see that β divides a_n. If f has degree 2 or 3, then f has a linear factor if and only if it is reducible over F. We record these observations as the first irreducibility test.

Proposition 5.1 (Rational Root Test) *Let R be a UFD with quotient field F. Suppose that $f(x) = a_0 + \cdots + a_n x^n \in R[x]$ has a root $\alpha/\beta \in F$ with $\gcd(\alpha, \beta) = 1$. Then α divides a_0, and β divides a_n. If $\deg(f) \le 3$, then f is irreducible over F if and only if f has no roots in F.*

Example 5.2 The polynomial $x^2 - p$ is irreducible over $\mathbb{Q}$ if p is a prime, as is $x^3 + 3x + 1$, by the rational root test. The cubic $x^3 + 2x^2 - 4x + 1$ factors as $(x-1)\left(x^2 + 3x - 1\right)$. The first factor could have been easily found by the rational root test; since $x^3 + 2x^2 - 4x + 1$ is monic, any rational root

of it is in $\mathbb{Z}$ and must divide 1, so ± 1 are the only possibilities. The fourth degree polynomial $x^4 - 4$ factors as $(x^2 - 2)(x^2 + 2)$, but it has no rational roots. Thus for polynomials of degree 4 or larger, the existence of a rational root is not necessary for a polynomial to factor over $\mathbb{Q}$.

The next irreducibility test is the one we use the most in this book.

Proposition 5.3 (Eisenstein Criterion) *Let R be a UFD with quotient field F, and let $f(x) = a_0 + \cdots + a_n x^n \in R[x]$.*

1. *Suppose there is a prime element $\pi \in R$ such that π divides $a_0, \ldots, a_{n-1}$ but not a_n and that π^2 does not divide a_0. Then f is irreducible over F.*
2. *Suppose there is a prime element $\pi \in R$ such that π divides $a_1, \ldots, a_n$ but not a_0 and that π^2 does not divide a_n. Then f is irreducible over F.*

Proof. We prove statement 1; the proof of statement 2 is similar. By factoring out the content of f, we may suppose that f is primitive. Consider the ring $R/(\pi)[x]$, and let $\overline{h}$ denote the image of $h \in R[x]$ obtained by reducing coefficients modulo π. The condition on the coefficients of f shows that $\overline{f} = \overline{a_n} x^n \neq 0$. Suppose that f factors over F. Then by Gauss' lemma, f also factors over R. Say $f = gh$ with $g, h \in R[x]$ both nonconstant polynomials. Then $\overline{f} = \overline{g} \cdot \overline{h}$ in $R/(\pi)[x]$ by Lemma 4.2. Since $\overline{f} = \overline{a_n} x^n$, then $\overline{g}$ and $\overline{h}$ each must be of the form cx^i, since the only monic irreducible factors of x^n are powers of x. If $\deg(\overline{g})$ and $\deg(\overline{h})$ are both positive, then π divides the constant term of both g and h. But a_0 is the product of these constant terms, which would force π^2 to divide a_0, which is false. Therefore, either $\overline{g}$ or $\overline{h}$ is a constant. Since g and h are nonconstant, π divides the leading coefficient of g or h. But a_n is the product of these leading coefficients, so π divides a_n, which is false. The only possibility left is that f is irreducible over F, which proves the criterion. □

Example 5.4 The polynomial $x^5 - 12x^3 + 2x + 2$ is irreducible over $\mathbb{Q}$ by an application of the Eisenstein criterion with $\pi = 2$. Similarly, with $\pi = 3$, the polynomial $x^3 - 3x + 3$ is irreducible. The Eisenstein criterion does not tell anything about $x^4 + 2x + 4$, since there is no prime that satisfies all the needed conditions.

Let p be a prime. The polynomial $x^p - 1$ factors as

$$x^p - 1 = (x - 1)(x^{p-1} + x^{p-2} + \cdots + x + 1).$$

The Eisenstein criterion would appear to be useless to determine whether $x^{p-1} + x^{p-2} + \cdots + x + 1$ is irreducible. However, this is not the case. By a change of variables, we can determine that this polynomial is irreducible

over $\mathbb{Q}$. Before doing so, however, we formalize the idea in the following lemma. The proof is straightforward and is left for the reader (Problem 12).

Lemma 5.5 *Let $f(x) \in R[x]$ and $a \in R$. Then the map $f(x) \mapsto f(x+a)$ is a ring isomorphism. Therefore, if $f(x+a)$ is irreducible, then $f(x)$ is irreducible.*

Example 5.6 Let $f(x) = x^p - 1$ with p a prime. Then

$$\begin{aligned} f(x+1) &= (x+1)^p - 1 \\ &= x^p + px^{p-1} + \binom{p}{2}x^{p-2} + \cdots + px. \end{aligned}$$

Since $x^p - 1$ factors as $(x-1)(x^{p-1} + \cdots + x + 1)$, replacing x by $x+1$ yields $f(x+1) = xg(x)$ with $g(x)$ the image of $x^{p-1} + \cdots + x + 1$ after substituting $x+1$ for x. Therefore, $x^{p-1} + \cdots + x + 1$ is irreducible if $g(x)$ is irreducible. However, the coefficients of $g(x) = x^{p-1} + px^{p-2} + \cdots + p$ are all binomial coefficients of the form $\binom{p}{i}$, which are all divisible by p (see Problem 13), except for the leading coefficient $\binom{p}{p} = 1$. Since the constant term is p, Eisenstein's criterion with $\pi = p$ shows that $g(x)$ is irreducible; hence, $x^{p-1} + \cdots + x + 1$ is irreducible.

The following result is our last irreducibility test.

Proposition 5.7 *Let R be a UFD with quotient field F, and let $f(x) \in R[x]$ be monic. If $\pi \in R$ is a prime element and if $\overline{f} \in R/(\pi)[x]$ is irreducible, then f is irreducible over F.*

Proof. The polynomial f is primitive since it is monic. If f factors over F, then f factors over R by Gauss' lemma. If $f = gh$ with $g, h \in R[x]$ nonconstant polynomials, then $\overline{f} = \overline{g} \cdot \overline{h}$ in $R/(\pi)[x]$ by Lemma 4.2. If $\overline{f}$ is irreducible, this means $\overline{g}$ or $\overline{h}$ is a unit in $R/(\pi)$. Since g and h are nonconstant, this would force π to divide the leading coefficient of either g or h, which cannot happen since f is monic. Therefore, f is irreducible over R, so f is also irreducible over F. □

The converse of this proposition is false, since $x^2 + x + 1$ is irreducible over $\mathbb{Z}$, but $x^2 + x + 1 = (x+2)(x+2)$ over $\mathbb{F}_3$. Also, if f is not monic, then the result is false, since $2x^2 + 3x + 1 = (2x+1)(x+1)$ factors over $\mathbb{Z}$, but its image in $\mathbb{F}_2[x]$ is $x+1$, which is irreducible.

Example 5.8 Over $\mathbb{F}_2$, the polynomials $1 + x^3 + x^4$ and $1 + x^3 + x^6$ can be seen to be irreducible by trial and error. Therefore, $1 + x^3 + x^4$ and $1 + x^3 + x^6$ are irreducible over $\mathbb{Q}$, as are $3 + 5x^3 + 7x^4$ and $-1 + 11x^3 + x^6$.

Example 5.9 Let p be a prime, and consider $x^p - x - 1$. This polynomial has no roots in $\mathbb{F}_p$, since every element of $\mathbb{F}_p$ is a root of $x^p - x$. While a polynomial can have no roots but be reducible, in this case this does not happen. Problem 3 of Section 10 shows that a prime degree polynomial that has no roots in a field F is irreducible over F under the following hypothesis: For any field K containing F, if the polynomial has a root in K, then it factors into linear factors in K. Using the result of this exercise, we show that the hypothesis holds for $x^p - x - 1$, which then implies that it is irreducible over $\mathbb{F}_p$, and so $x^p - x - 1$ is irreducible over $\mathbb{Q}$.

Suppose that K is a field containing $\mathbb{F}_p$ for which $x^p - x - 1$ has a root a. So $a^p - a = 1$. We claim that $a + 1, a + 2, \dots, a + p - 1$ are also roots of $x^p - x - 1$ in K. To see this, if $1 \le i \le p - 1$, then $(a + i)^p - (a + i) - 1 = a^p + i^p - a - i - 1 = a^p - a - 1 = 0$, since $i^p = i (\bmod p)$ by Fermat's little theorem. Therefore, we have p roots of $x^p - x - 1$ in K, so $x^p - x - 1$ factors into linear factors in K. Therefore, Problem 3 of Section 10 shows that $x^p - x - 1$ is irreducible over $\mathbb{F}_p$, since it has no root in $\mathbb{F}_p$.

Problems

1. For any positive integer n, give an example of a ring of characteristic n.

2. Let F be a field. If $\text{char}(F) = p > 0$, show that the prime subring of R is isomorphic to the field $\mathbb{F}_p$, and if $\text{char}(F) = 0$, then the prime subring is isomorphic to $\mathbb{Z}$.

3. Let F be a field. The *prime subfield of* F is the intersection of all subfields of F. Show that this subfield is the quotient field of the prime subring of F, is contained inside all subfields of F, and is isomorphic to $\mathbb{F}_p$ or $\mathbb{Q}$ depending on whether the characteristic of F is $p > 0$ or 0.

4. Let R be an integral domain. Show that a and b are associates in R if and only if $aR = bR$.

5. Show that 2, 3, and $1 \pm \sqrt{-5}$ are all irreducible in the ring
$$\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : a, b \in \mathbb{Z}\}$$
and $2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$, but that 2 does not divide either of $1 \pm \sqrt{-5}$.

6. If R is a UFD, show that $\pi_1^{\min\{e_1, f_1\}} \cdots \pi_n^{\min\{e_n, f_n\}}$ is a gcd of $\pi_1^{e_1} \cdots \pi_n^{e_n}$ and $\pi_1^{f_1} \cdots \pi_n^{f_n}$.

7. Let R be a ring, and let S be a subset of R that is closed under multiplication and does not contain 0. Use a Zorn's lemma argument to show that there are ideals of R maximal with respect to being disjoint to S. Also show that any such ideal is prime by mimicking the argument used in the proof of Theorem 2.3.

8. If R is an integral domain such that there are primes $\pi_1, \ldots, \pi_n$, $\theta_1, \ldots, \theta_m$ with $\pi_1 \cdots \pi_n = \theta_1 \cdots \theta_m$, show that $m = n$, and after renumbering if necessary, show that π_i and θ_i are associates for each i.

9. Use the previous problem to show that an integral domain R is a UFD if and only if every nonunit in R can be factored into a product of primes.

10. Show that an integral domain R is a UFD if and only if (1) every nonunit of R can be factored into a product of irreducibles and (2) every irreducible is prime.

11. Let $R = \mathbb{Q}[x^2, x^3]$, the set of all polynomials over $\mathbb{Q}$ with no x term. Show that a gcd of x^5 and x^6 does not exist in R.

12. Let R be a ring. If $a \in R$, show that the map $f(x) \mapsto f(x+a)$ is a ring isomorphism of $R[x]$.

13. If p is a prime, show that p divides the binomial coefficient $\binom{p}{i}$ if $0 < i < p$.

14. Let $X = \{x_\alpha\}_{\alpha \in I}$ be a set of variables and $F[X]$ the ring of polynomials in the variables from X. This ring can be thought of as the union of all the rings $F[x_{\alpha_1}, \ldots, x_{\alpha_n}]$, as the union ranges over all finite subsets $\{x_{\alpha_1}, \ldots, x_{\alpha_n}\}$ of X. Show that $F[X]$ is a UFD if F is a field.

15. Let $R = \mathbb{Z}[x]$, a UFD. Show that 2 and x are relatively prime but that 1 is not a linear combination of 2 and x; that is, there are no elements $f, g \in \mathbb{Z}[x]$ with $1 = 2f + xg$.

16. Let R be a UFD and $a_1, \ldots, a_n \in R$. Prove that

$$\gcd(a_1, \ldots, a_n) = \gcd(a_1, \gcd(a_2, \ldots, a_n)).$$

Conclude that a gcd of any finite set exists.

17. Let R be a PID. If $d = \gcd(a, b)$, show that $dR = aR + bR$, and conclude that d is a linear combination of a and b.

18. This problem fills in all the details of the proof that $R[x]$ is a UFD if R is a UFD. Some of these parts are standard facts of localization but are included in case the reader has not seen localization beyond the construction of the field of quotients of an integral domain.

(a) Let $A \subseteq B$ be commutative rings, and let Q be a prime ideal of B. Show that $Q \cap A$ is a prime ideal of A.

(b) If $A \subseteq B$, suppose that there is a subset S of A that is closed under multiplication, every element of S is a unit in B, and $B = \{a/s : a \in A, s \in S\}$. If $a \in A - S$, show that $aB \cap A = aA$. We write $B = A_S$ when B is of this form.

(c) Let $A \subseteq B$, and suppose that there is a set S as in Problem 18b with $B = A_S$. If P is a prime ideal of A with $P \cap S = \varnothing$, show that PB is a prime ideal of B and that $PB \cap A = P$.

(d) If R is an integral domain with quotient field F, and if $S = R - \{0\}$, show that $F[x] = R[x]_S$.

(e) Put the previous steps together to prove Theorem 4.5 in full detail.

Appendix B
Set Theory

In this appendix, we discuss Zorn's lemma and cardinal arithmetic. For more information on these topics, see Enderton [8] or Stoll [26].

1 Zorn's Lemma

In this book, we use Zorn's lemma in algebra to prove the isomorphism extension theorem, the existence of an algebraic closure, and some other results. We point out that Zorn's lemma has a large number of equivalent formulations; for instance, Zorn's lemma is equivalent to the axiom of choice and to the well ordering principle. However, we only require the statement of Zorn's lemma in this book.

We now describe the terms involved in the statement of Zorn's lemma. A *partial order* $\leq$ on a set S is a binary relation such that (1) $s \leq s$ for all $s \in S$, (2) if $s \leq t$ and $t \leq s$, then $s = t$, and (3) if $r \leq s$ and $s \leq t$, then $r \leq t$. Examples of a set with a partial order include the real numbers with the usual ordering, and the set of all subsets of a given set, with set inclusion as the order. If S is a set with partial order $\leq$, we shall refer to the pair $(S, \leq)$ as a *partially ordered set.*

Let $(S, \leq)$ be a partially ordered set. An element $m \in S$ is said to be *maximal* if whenever $s \in S$ with $m \leq s$, then $s = m$. If T is a subset of S, then an element $s \in S$ is said to be an *upper bound* for T if $t \leq s$ for all $t \in T$. For instance, if S is the set of all subsets of $\{1, 2, 3, 4\}$, then $\{1, 2, 3, 4\}$ is a maximal element of S. If T is the set of all proper subsets of

$\{1, 2, 3, 4\}$, then $\{1, 2, 3, 4\}$ is an upper bound for T. Note that this upper bound is not in T. Also, $\{1, 2, 3\}$ and $\{1, 2, 4\}$ are both maximal elements of T. Finally, a subset T of a partially ordered set $(S, \leq)$ is said to be a *chain* if for every $t_1, t_2 \in T$, then either $t_1 \leq t_2$ or $t_2 \leq t_1$. With the example above, $\{\varnothing, \{1\}, \{1, 2\}, \{1, 2, 4\}\}$ is a chain in S.

We can now state Zorn's lemma.

Theorem 1.1 (Zorn's Lemma) *Let $(S, \leq)$ be a nonempty partially ordered set. Suppose that for any chain T in S there is an upper bound for T in S. Then S contains a maximal element.*

In the statement of Zorn's lemma, an upper bound for a chain T need not be an element of T, merely an element of S.

Example 1.2 Here is the first place that Zorn's lemma usually arises in algebra. Let R be a ring with identity. We show that R contains a maximal ideal. Let $\mathcal{S}$ be the set of all proper ideals of R. Then $\mathcal{S} \neq \varnothing$, since $(0) \in \mathcal{S}$. The set $\mathcal{S}$ is partially ordered by set inclusion. To verify that Zorn's lemma applies, let T be a chain in $\mathcal{S}$. Define I to be $\bigcup T$, the union of all ideals in T. We can see that I is an ideal of R, for if $a, b \in I$, then $a, b \in J$ for some $J \in T$, since T is a chain. Then $a - b \in J \subseteq I$. Also, if $a \in I$ and $r \in R$, then $a \in J$ for some $J \in T$, so $ra, ar \in J \subseteq I$. Thus, I is an ideal of R. Moreover, I is a proper ideal of R since no $J \in T$ contains 1, so I does not contain 1. Therefore, $I \in \mathcal{S}$. By Zorn's lemma, $\mathcal{S}$ contains a maximal element M. A maximal ideal of R is precisely a maximal element of the set of proper ideals of R, so M is a maximal ideal of R.

We now give a couple of general examples of how Zorn's lemma can be used in algebra. All of the uses of Zorn's lemma in this book, including the example above, are special examples of these. Appendix D uses Zorn's lemma to prove that any vector space contains a basis.

Example 1.3 Let X be a set, and let S be a nonempty collection of subsets of X, with the partial order of set inclusion. Suppose that for every chain T in S the set $\bigcup T$ is an element of S. Then S has a maximal element. To verify this, all we need to see to apply Zorn's lemma is that the chain T has an upper bound in S. But the union $\bigcup T$ clearly is an upper bound for T, since any $t \in T$ is a subset of this union. The assumption is that this union is in S; hence, Zorn's lemma applies.

Example 1.4 Let X and Y be sets, and let S be a nonempty collection of pairs (A, f), where A is a subset of X and $f : A \to Y$ is a function. We can define a partial order on S as follows: Let $(A, f) \leq (B, g)$ if $A \subseteq B$ and $g|_A = f$. It is easy to see that $\leq$ is indeed a partial order on S. Suppose that T is a chain in S. Let $M = \bigcup T$, and define a function $h : T \to Y$ by $h(x) = g(x)$ if $(X, g) \in T$ and $x \in X$. The function h is well defined by the

condition that T is a chain. Suppose that for each chain T, the pair (M, h) as constructed is an element of S. Then S has a maximal element. This follows from Zorn's lemma because the element (M, h) is an upper bound for T by construction and, by hypothesis, lies in S.

2 Cardinality and Cardinal Arithmetic

We will require the use of cardinal arithmetic in a couple of places in this book. The theorem that any two bases of a finite dimensional vector space have the same number of elements can be extended to arbitrary vector spaces by using Zorn's lemma and some results of cardinal arithmetic. We now give the basic definitions and results on cardinal arithmetic that we require in this book.

If S and T are sets, we write $S \preceq T$ if there is an injective function from S to T. It is proved in most set theory texts that $S \preceq T$ if and only if there is a surjective function from T to S. If $S \preceq T$ and $T \preceq S$, then we say S and T have the same cardinality and write $S \approx T$. The *Schröder–Bernstein* theorem says that this is equivalent to the existence of a bijection between S and T. We will write $S \prec T$ if $S \preceq T$ and if S and T do not have the same cardinality.

The cardinality of a set S will be denoted $|S|$. Addition and multiplication of cardinal numbers is defined by $|S| + |T| = |S \uplus T|$, where $S \uplus T$ is the disjoint union of S and T. Also, $|S| \cdot |T| = |S \times T|$. We write $|S| \leq |T|$ and $|S| < |T|$ if $S \preceq T$ and $S \prec T$, respectively. If S is an infinite set, then $|S|$ is called an *infinite cardinal.* If S is finite or if $S \approx \mathbb{N}$, then S is said to be *countable*. If S is countable and infinite, we write $|S| = \aleph_0$. The cardinal $\aleph_0$ is the smallest infinite cardinal; that is, if S is a countably infinite set and T is any infinite set, there is an injective function $S \to T$. We recall the basic facts of cardinal arithmetic in the following proposition.

Proposition 2.1 *Let S and T be sets.*

1. *If T is infinite and if $\{S_n : n \in \mathbb{N}\}$ is a collection of subsets of S with $|S_n| \leq |T|$ for all n, then $\left|\bigcup_{n\in\mathbb{N}} S_n\right| \leq |T|$.*
2. *If S and T are sets, then $|S| \leq |S| + |T|$. If either S or T is infinite, then $|S| + |T| = \max\{|S|, |T|\}$.*
3. *If S and T are nonempty sets, then $|S| \leq |S| \cdot |T|$. If either S or T is infinite, then $|S| \cdot |T| = \max\{|S|, |T|\}$.*
4. *If T is an infinite set, then $\aleph_0 \cdot |T| = |T|$.*

Example 2.2 Let X be a set, and let $\mathcal{P}(X)$ be the set of all subsets of X. We show that $|\mathcal{P}(X)| > |X|$. Note that there is an injective map $X \to \mathcal{P}(X)$

given by $a \mapsto \{a\}$. Therefore, $|X| \le |\mathcal{P}(X)|$. We finish the proof by showing that there is no surjective map from X to $\mathcal{P}(X)$. Let $f : X \to \mathcal{P}(X)$ be any function. Define S by $S = \{a \in X : a \notin f(a)\}$. We claim that S is not in the image of f. Suppose instead that $S = f(x)$ for some x. Then $x \in S$ if and only if $x \notin f(x) = S$. This is impossible, so $S \notin \text{im}(f)$.

Problems

1. Use Zorn's lemma to prove that if R is a ring with identity, then R has a maximal left ideal.

2. In this problem, we show that a ring without an identity may not have any maximal ideals. Let p be a prime, and let $R = \{a/p^n : a \in \mathbb{Z}, n \ge 0\}$. Then R is a subgroup of $\mathbb{Q}$ under addition. Define multiplication in R by $x \cdot y = 0$ for all $x, y \in R$. Note that a subset of R is an ideal if and only if it is a subgroup under addition. Show that the only subgroups of R are the cyclic subgroups generated by $1/p^n$ for some n, and conclude that R does not have a maximal ideal.

3. Let R be a commutative ring, and let S be a subset of R that is closed under multiplication and does not contain 0. Use Zorn's lemma to show that there is an ideal maximal with respect to being disjoint from S.
   (This fact is used to prove that the intersection of all prime ideals containing an ideal I is equal to the radical of I.)

4. Prove the Schröder–Bernstein theorem.

5. Prove that $\mathbb{Z} \simeq \mathbb{N}$ and $\mathbb{Q} \simeq \mathbb{N}$.

6. Prove that $\mathbb{N} \prec \mathbb{R}$.

Appendix C
Group Theory

There are a number of results from group theory that we will need in Galois theory. This section gives a brief survey of these results. For a more complete treatment of group theory, see Rotman [23] or any of the general algebra texts.

1 Fundamentals of Finite Groups

Let G be a group, and let H be a subgroup of G. Recall that the *left coset* gH of an element $g \in G$ is the set of all elements of the form gh with $h \in H$. Right cosets are defined similarly. The distinct left (or right) cosets of H partition G. If G is finite, then each coset has the same number of elements. These facts form the heart of the proof of Lagrange's theorem, the most fundamental result about finite groups.

Theorem 1.1 (Lagrange) *If H is a subgroup of a finite group G, then $|H|$ divides $|G|$. Moreover, if $[G : H]$ is the number of cosets of H in G, then $|G| = |H| \cdot [G : H]$.*

Proof. The proof of the first statement can be found in any book on group theory. Lagrange's theorem usually is stated as just the first sentence. The proof yields the equality $|G| = |H| \cdot [G : H]$. □

If G is a group and if N is a subgroup of G, then N is said to be a *normal subgroup* of G if $gng^{-1} \in N$ for all $g \in G$ and $n \in N$. If N is

a normal subgroup of G, let G/N be the set of all left cosets of N in G. Then G/N can be given the structure of a group by defining multiplication by $gN \cdot hN = ghN$. This definition is well defined, independent of the representation of the cosets.

Suppose that G is a finite Abelian group. Then there is a complete description of the structure of G. The following theorem is often called the *fundamental theorem of finite Abelian groups.*

Theorem 1.2 (Fundamental Theorem of Finite Abelian Groups) *Let G be a finite Abelian group. Then G is a direct product of cyclic subgroups. Therefore, $G \cong \mathbb{Z}/n_1\mathbb{Z} \times \cdots \times \mathbb{Z}/n_r\mathbb{Z}$ for some integers n_i.*

It is not hard to show that $\mathbb{Z}/nm\mathbb{Z} \cong \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$ if $\gcd(n, m) = 1$. This fact is one formulation of the Chinese remainder theorem. From this fact and the fundamental theorem of finite Abelian groups, one can obtain the following description of finite Abelian groups.

Corollary 1.3 *Let G be a finite Abelian group.*

1. *There are integers $n_1, \ldots, n_r$, where n_i divides n_{i-1} for each i, such that $G \cong \mathbb{Z}/n_1\mathbb{Z} \times \cdots \times \mathbb{Z}/n_r\mathbb{Z}$. The n_i are uniquely determined by G and are called the invariant factors of G.*

2. *There are integers m_{ij} and primes p_i such that $G \cong \mathbb{Z}/p_1^{m_{11}}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_t^{m_{tu}}\mathbb{Z}$. The various $p_i^{m_{ij}}$ are uniquely determined by G and are called the elementary divisors of G.*

Let G be a finite group. Then the *exponent* of G, denoted $\exp(G)$, is the least common multiple of the orders of the elements of G. For example, the exponent of the symmetric group S_3 is 6. We give a useful result about the exponent of a finite Abelian group.

Proposition 1.4 *Let G be a finite Abelian group. If $n = \exp(G)$, then there is an element of G of order n. Therefore, $\exp(G)$ is the maximum order of an element of G. Furthermore, G is cyclic if and only if $|G| = \exp(G)$.*

Proof. A short calculation using the decomposition of G into a product of cyclic groups shows that for every divisor m of $|G|$ there is an element of order m. If $n = \exp(G)$, then n divides $|G|$ by Lagrange's theorem and the definition of least common multiple. Therefore, G contains an element of order n. Since a group G is cyclic if and only if it contains an element whose order is $|G|$, we see that G is cyclic if and only if $|G| = \exp(G)$. □

An alternative proof of this proposition that does not invoke the fundamental theorem of finite Abelian groups is outlined in Problem 1.

2 The Sylow Theorems

Let G be a finite group, and let p be a prime dividing the order $|G|$ of G. Let $|G| = p^n q$ with q not divisible by p. A *p-Sylow subgroup* of G is a subgroup of order p^n, the maximal power of p possible for a subgroup of G. The Sylow theorems give existence and properties of p-Sylow subgroups of a finite group.

Theorem 2.1 (First Sylow Theorem) *Let G be a finite group, and let p be a prime divisor of $|G|$. Then there exists a p-Sylow subgroup of G.*

Theorem 2.2 (Second Sylow Theorem) *Let p be a prime divisor of $|G|$. If H is a subgroup of G of order a power of p, then $H \subseteq xPx^{-1}$ for some p-Sylow subgroup P of G. In particular, if P_1 and P_2 are two p-Sylow subgroups of G, then $P_2 = xP_2x^{-1}$ for some $x \in G$.*

Theorem 2.3 (Third Sylow Theorem) *Let p be a prime divisor of $|G|$. If n is the number of p-Sylow subgroups of G, then p divides $|G|$ and $n \equiv 1 (\bmod p)$.*

The first Sylow theorem is the best partial converse of Lagrange's theorem. Given a divisor m of $|G|$, there need not be a subgroup of G of order m. For instance, there is no subgroup of the alternating group A_4 of order 6. However, if $|G| = p^n q$ as above and if $m = p^n$, then the first Sylow theorem gives the existence of a subgroup of order m.

Some of the power of the Sylow theorems comes from the following two facts. First, it is often convenient to have a subgroup H of a group G with $|H|$ and $[G : H]$ relatively prime, as is the case if H is a Sylow subgroup. Second, groups of prime power order are very nicely behaved. We shall see one property of such groups shortly. If G is a group of order p^n with p a prime, then G is said to be a *p-group*. If G is an arbitrary group, a subgroup H of G is said to be a *maximal subgroup* of G if H is a proper subgroup of G that is not contained in any subgroup of G other than G and itself. The following result will help to use p-groups in field theory, for instance, in the proof of the fundamental theorem of algebra in Section 5. An outline of a proof of this proposition can be found in Problem 2.

Proposition 2.4 *Let G be a p-group of order p^n. If H is a maximal subgroup of G, then H is normal in G and $[G : H] = p$.*

If G is a finite group, then maximal subgroups of G always exist. Using this proposition repeatedly, we can extend the first Sylow theorem.

Corollary 2.5 *Let G be a group of order $p^n q$ with p a prime. Then G contains a subgroup of order p^r for any $r \leq n$.*

3 Solvable Groups

In many ways, abstract algebra began with the work of Abel and Galois on the solvability of polynomial equations by radicals. The key idea Galois had was to transform questions about fields and polynomials into questions about finite groups. For the proof that it is not always possible to express the roots of a polynomial equation in terms of the coefficients of the polynomial using arithmetic expressions and taking roots of elements, the appropriate group theoretic property that arises is the idea of solvability.

Definition 3.1 *A group G is solvable if there is a chain of subgroups*

$$\langle e \rangle = H_0 \subseteq H_1 \subseteq \cdots \subseteq H_n = G$$

such that, for each i, the subgroup H_i is normal in H_{i+1} and the quotient group H_{i+1}/H_i is Abelian.

An Abelian group G is solvable; the chain of subgroups $\langle e \rangle \subset G$ satisfies the definition. Also, the symmetric groups S_3 and S_4 are solvable by considering the chains $\langle e \rangle \subset A_3 \subset S_3$ and $\langle e \rangle \subset H \subset A_4 \subset S_4$, respectively, where

$$H = \{e, (12)(34), (13)(24), (14)(23)\}.$$

Likewise, any p-group is solvable, since if $|G| = p^n$, there is a chain of subgroups

$$\langle e \rangle \subset N_1 \subset N_2 \subset \cdots \subset N_n = G$$

where $|N_i| = p^i$ and N_{i-i} is normal in N_i, by Proposition 2.4. Thus, N_i/N_{i-1} has order p; hence, it is cyclic and therefore Abelian. One can obtain such a chain by taking N_{n-1} to be any maximal subgroup of G, N_{n-2} a maximal subgroup of N_{n-1}, and so on, and using Proposition 2.4. We shall show below that S_n is not solvable if $n \geq 5$. This is the group theoretic result we need to show that the roots of the general polynomial of degree n cannot be written in terms of the coefficients of the polynomial by using algebraic operations and extraction of roots.

We now begin to work toward showing that the symmetric group S_n is not solvable if $n \geq 5$. If G is a group, let G' be the *commutator subgroup* of G; that is, G' is the subgroup of G generated by all $ghg^{-1}h^{-1}$ with $g, h \in G$. It is an easy exercise to show that G' is a normal subgroup of G and that G/G' is Abelian. In fact, if N is a normal subgroup of G, then G/N is Abelian if and only if $G' \subseteq N$. We define $G^{(i)}$ by recursion by setting by $G^{(1)} = G'$ and $G^{(i+1)} = (G^{(i)})'$. We then obtain a chain

$$G \supseteq G^{(1)} \supseteq G^{(2)} \supseteq \cdots \supseteq G^{(n)} \supseteq \cdots$$

such that $G^{(m+1)}$ is normal in $G^{(m)}$ and $G^{(m)}/G^{(m+1)}$ is Abelian for all m.

Lemma 3.2 *G is solvable if and only if $G^{(n)} = \langle e \rangle$ for some n.*

Proof. Suppose that $G^{(n)} = \langle e \rangle$ for some n. Then the chain

$$G \supseteq G^{(1)} \supseteq \cdots \supseteq G^{(n)} = \langle e \rangle$$

shows that G is solvable. Conversely, suppose that G is solvable, and let

$$\langle e \rangle = H_n \subset H_{n-1} \subset \cdots \subset H_0 = G$$

be a chain of subgroups such that H_{m+1} normal in H_m and H_m/H_{m+1} is Abelian for all m. Then G/H_1 is Abelian, so $G' = G^{(1)} \subseteq H_1$. Thus, $(G^{(1)})' \subseteq H_1'$. Because H_1/H_2 is Abelian, $H_1' \subseteq H_2$. Therefore, $G^{(2)} = (G^{(1)})' \subseteq H_2$. Continuing this process shows that $G^{(n)} \subseteq H_n = \langle e \rangle$, so $G^{(n)} = \langle e \rangle$. □

Proposition 3.3 *Let G be a group, and let N be a normal subgroup of G. Then G is solvable if and only if N and G/N are solvable.*

Proof. We have $N^{(m)} \subseteq G^{(m)}$ and $(G/N)^{(m)} = (G^{(m)}N)/N$ for all m. Thus, if G is solvable, there is an n with $G^{(n)} = \langle e \rangle$. Therefore, $N^{(n)} = \langle e \rangle$ and $(G/N)^{(n)} = \langle e \rangle$, so both N and G/N are solvable. Conversely, suppose that N and G/N are solvable. Then there is an m with $(G/N)^{(m)} = \langle e \rangle$, so $G^{(m)} \subseteq N$. There is an n with $N^{(n)} = \langle e \rangle$, so $G^{(n+m)} = (G^{(m)})^{(n)} \subseteq N^{(n)} = \langle e \rangle$. Therefore, $G^{(n+m)} = \langle e \rangle$, so G is solvable. □

Lemma 3.4 *If $n \geq 5$, then A_n is a simple group.*

For a proof of this important result, see Hungerford [13, p. 49].

Corollary 3.5 *If $n \geq 5$, then S_n is not solvable.*

Proof. Since A_n is simple and non-Abelian, $A_n' = A_n$. Thus, we see for all m that $A_n^{(m)} = A_n \neq \langle e \rangle$, so A_n is not solvable. By the proposition above, S_n is also not solvable. □

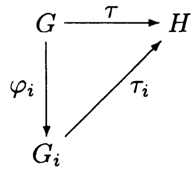
4 Profinite Groups

We give here a brief description of profinite groups. These are the groups that arise as the Galois group of a Galois extension of any degree, possibly infinite. This information is only used in Sections 17 and 18. Most of the results are stated without proof. The interested reader can find proofs and more information about profinite groups in Serre [24] and Shatz [25].

Let $\{G_i\}_{i \in I}$ be a collection of groups. Suppose that I is a *directed set*. This means that I has a partial order $\leq$ such that for any $i, j \in I$, there is a $k \in I$ with $i \leq k$ and $j \leq k$. Suppose that for each i and j with $i \leq j$

there is a group homomorphism $\varphi_{i,j} : G_j \to G_i$. Moreover, suppose that whenever $i \le j \le k$ we have $\varphi_{i,k} = \varphi_{j,k} \circ \varphi_{i,j}$. Then the set of groups $\{G_i\}$ together with the homomorphisms $\varphi_{i,j}$ are said to form an *inverse system* of groups.

Definition 4.1 *Let $\{G_i, \varphi_{i,j}\}$ be an inverse system of groups. The inverse limit of this system is a group G together with homomorphisms $\varphi_i : G \to G_i$ such that if $i \le j$, then $\varphi_i = \varphi_{i,j} \circ \varphi_j$, along with the following universal mapping property: If H is a group together with homomorphisms $\tau_i : H \to G_i$ such that $\tau_i = \varphi_{i,j} \circ \tau_j$ whenever $i \le j$, then there is a unique group homomorphism $\tau : H \to G$ with $\tau_i = \varphi_i \circ \tau$ for each i; that is, the following diagram commutes:*



The following proposition shows that inverse limits exist and are unique up to isomorphism.

Proposition 4.2 *Let $\{G_i, \varphi_{i,j}\}$ be an inverse system of groups. Then the inverse limit of the G_i exists and is unique up to isomorphism.*

Proof. Let $\prod_i G_i$ be the dirtct product of the G_i. Define G by

$$G = \left\{ \{g_i\} \in \prod_i G_i : \varphi_{i,j}(g_j) = g_i \text{ for each pair } (i,j) \text{ with } i \le j \right\}.$$

Then G is a subgroup of $\prod_i G_i$, since the $\varphi_{i,j}$ are homomorphisms. Let $\varphi_i : G \to G_i$ be the restriction to G of the usual projection map. If $i \le j$, then $\varphi_i = \varphi_{i,j} \circ \varphi_j$ by the definition of G. To verify the universal mapping property, let H be a group with homomorphisms $\tau_i : H \to G_i$ such that $\tau_i = \varphi_{i,j} \circ \tau_j$ whenever $i \le j$. Define a homomorphism $\tau : H \to \prod_i G_i$ by $\tau(h) = \{\tau_i(h)\}$. The condition $\tau_i = \varphi_{i,j} \circ \tau_j$ says precisely that $\mathrm{im}(\tau) \subseteq G$. Thus, τ is a homomorphism from H to G. The formula for τ is forced upon us by the requirement that $\tau_i = \varphi_i \circ \tau$, so τ is unique. Thus, G is an inverse limit of the G_i. □

We can now define a profinite group.

Definition 4.3 *A profinite group is an inverse limit of finite groups.*

There is a natural topology on a profinite group. If $\{G_i\}$ is an inverse system of finite groups, give each G_i the discrete topology and then give $\prod_i G_i$

the product topology. The inverse limit of the G_i then inherits the subspace topology from $\prod_i G_i$. This topology is an important tool for studying profinite groups and is used frequently in proofs of the results stated in this section. We describe a relation between the topology and the algebra of G. Let $N_i = \ker(\varphi_i)$. Then G/N_i is isomorphic to a subgroup of G_i; consequently, N_i is a normal subgroup of finite index. Moreover, since $N_i = \varphi_i^{-1}\{0\}$, the preimage of a single point, N_i is both open and closed, since G_i has the discrete topology.

Proposition 4.4 *Let G be a profinite group. As a topological space, G is Hausdorff, compact, and totally disconnected.*

Many of the fundamental numerical results about finite groups have analogs in the theory of profinite groups. First, we need a meaningful definition of the order of a profinite group. A *supernatural number* is a formal product $\prod_p p^{n_p}$, where p runs over all primes, and the exponents are nonnegative integers or ∞. While there is no natural way to add supernatural numbers, the product, greatest common divisor, and least common multiple of a set of supernatural numbers can be defined in the obvious way. By using supernatural numbers, we can give a useful definition of the order of a group and the index of a subgroup.

Definition 4.5 *Let G be the inverse limit of the finite groups $\{G_i\}$.*

1. *The order of G is the supernatural number* $\operatorname{lcm}_i \{|G_i|\}$.
2. *If H is a closed subgroup of G, then the index $[G : H]$ is equal to* $\operatorname{lcm}_i \{[G_i : G_i \cap H]\}$.

If p is a prime and n_i is the power of p occurring in $|G_i|$, then $\max\{n_i\}$ is the power of p occurring in $|G|$. Even though each n_i is finite, the maximum may be infinite. This is the reason for allowing an exponent of ∞ in a supernatural number.

We record the basic numerical properties of profinite groups. The first part of the following proposition is an analog of Lagrange's theorem.

Proposition 4.6 *Let G be a profinite group.*

1. *If $H \subseteq K$ are closed subgroups of G, then $[G : K] = [G : H] \cdot [H : K]$.*
2. *If H is a closed subgroup of G, then $[G : H] = \operatorname{lcm}_U \{[G/U : HU/U]\}$, where U ranges over all open normal subgroups U of G. In particular, $|G| = \operatorname{lcm}_U \{|G/U|\}$.*

Two different inverse systems of groups may have the same inverse limit. Part 2 of this proposition shows that indices are not dependent on a specific choice of inverse system.

There are good extensions of the Sylow theorems to the class of profinite groups. Let p be a prime. A *pro-p-group* is a profinite group G for which $|G| = p^n$ for some n with $1 < n \leq \infty$. Equivalently, a pro-p-group is an inverse limit of p-groups. Suppose that G is a profinite group whose order is divisible by a prime p. This means that $|G| = \prod_q q^{n_q}$, such that $n_p \geq 1$. A subgroup H of G is called a p-Sylow subgroup of G provided that H is a pro-p-group and $[G : H]$ is not divisible by p.

Theorem 4.7 *Let G be a profinite group, and let p be a prime divisor of $|G|$.*

1. *The group G has a p-Sylow subgroup.*
2. *If P is a pro-p-subgroup of G, then P is contained in a p-Sylow subgroup of G.*
3. *Any two p-Sylow subgroups of G are conjugate.*

Problems

1. This problem outlines a proof of Proposition 1.4 that does not use the fundamental theorem of finite Abelian groups. Let G be a finite Abelian group.

   (a) If $a, b \in G$ have orders n and m, respectively, show that the order of ab is nm if n and m are relatively prime.

   (b) If a has order n and if t is a divisor of n, show that a^t has order n/t.

   (c) If n and m are positive integers, and if G contains elements of order n and m, show that G contains an element of order $\operatorname{lcm}(n, m)$. Use this fact to prove that G contains an element of order $\exp(G)$.
   (Hint: factor $\operatorname{lcm}(n, m)$ into prime powers, and then use the first two parts of this problem.)

2. Let p be a prime and G be a p-group.

   (a) If H is a subgroup of G, show that $H \subset N(H)$, where $N(H) = \{g \in G : gHg^{-1} = H\}$.

   (b) If H is a maximal subgroup of G, show that H is normal in G and that $[G : H] = p$.

   (Hint: Recall that $Z(G) \neq \langle e \rangle$ if G is a p-group. Find a subgroup Z of $Z(G)$ of order p, consider G/Z, and use induction on n, where $|G| = p^n$.)

3. Define multiplication, greatest common divisor, and least common multiple of a set of supernatural numbers.

4. Let G be a profinite group, and let H be a closed subgroup of G. Show that $[G : H] = \mathrm{lcm}_U \{[G : U]\}$, as the U range over all open normal subgroups of G that contain U.

5. Let G be a profinite group, and let H be a subgroup of G. Show that the closure $\overline{H}$ of H is given by $\overline{H} = \bigcap_U HU$, as U ranges over all open normal subgroups of G.

6. Let G be a profinite group, and let H be a subgroup of G. Show that $\overline{H}$ is the intersection of all open normal subgroups of G containing H.

7. Let G be a profinite group. If N is a normal subgroup of finite index in G, show that N is open.

8. Let G be a profinite group. Show that any closed subgroup of G is also a profinite group. Also, show that any quotient of G is a profinite group.

9. Read Chapter I of Shatz [25] and prove the results on profinite groups stated in this section.

Appendix D
Vector Spaces

The use of the theory of vector spaces is a key element in field theory. In this appendix, we review the concepts that we will need. For a more detailed account of vector spaces, see Herstein [12] or Walker [27].

1 Bases and Dimension

The most important property of vector spaces is the existence of a basis. Let V be a vector space over a field F. If $v_1, \ldots, v_n \in V$, any element of the form $\alpha_1 v_1 + \cdots + \alpha_n v_n$ is called a *linear combination* of the v_i. A subset $\mathcal{B}$ of V is said to be *linearly independent* over F provided that whenever $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$, with $\alpha_i \in F$ and $v_i \in \mathcal{B}$, then each $\alpha_i = 0$. Therefore, $\mathcal{B}$ is linearly independent, provided that the only way to write 0 as a linear combination of elements of $\mathcal{B}$ is in the trivial way, where all coefficients are 0. If a set is not linearly independent, it is said to be *linearly dependent*. For example, any singleton set $\mathcal{B} = \{v\}$ with $v \neq 0$ is linearly independent. By definition, the empty set $\varnothing$ is linearly independent. Any set containing 0 is dependent.

If $\mathcal{B}$ is a subset of V, then $\mathcal{B}$ is said to *span* V if every element of V is a linear combination of elements of $\mathcal{B}$. For example, if $V = F^n$, the set of all n-tuples of elements of F, then the set

$$\{(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)\}$$

spans F^n. The set F^n also spans F^n. In fact, if $\mathcal{B}$ spans a vector space V, then any set containing $\mathcal{B}$ also spans V.

We can now define a basis. If V is an F-vector space, a set $\mathcal{B}$ is a *basis* for V if $\mathcal{B}$ is linearly independent and spans V. For finitely generated vector spaces (i.e., those spaces that are spanned by a finite set), proofs of the existence of a basis are standard. However, a use of Zorn's lemma shows that any vector space has a basis. Because this proof is less standard, we give it here. Moreover, this proof is a good example of how Zorn's lemma is used in algebra.

Theorem 1.1 *Let V be a vector space over a field F.*

1. *There exists a basis for V.*
2. *If $\mathcal{C}$ is any linearly independent set in V, then $\mathcal{C}$ is contained in a basis of V.*
3. *If $\mathcal{D}$ is any spanning set for V, then $\mathcal{D}$ contains a basis of V.*
4. *If $\mathcal{C} \subseteq \mathcal{D}$ are subsets of V such that $\mathcal{C}$ is linearly independent and $\mathcal{D}$ spans V, then there is a basis $\mathcal{B}$ with $\mathcal{C} \subseteq \mathcal{B} \subseteq \mathcal{D}$.*

Proof. We give a proof for part 4. Parts 2 and 3 follow from part 4 by setting $\mathcal{C} = \varnothing$ and $\mathcal{D} = V$, respectively. Part 1 follows from part 4 by setting $\mathcal{C} = \varnothing$ and $\mathcal{D} = V$. Suppose that $\mathcal{C} \subseteq \mathcal{D}$ such that $\mathcal{C}$ is linearly independent and $\mathcal{D}$ spans V. Let

$$\mathcal{S} = \{\mathcal{E} : \mathcal{C} \subseteq \mathcal{E} \subseteq \mathcal{D}, \ \ \mathcal{E} \text{ is linearly independent}\}.$$

The set $\mathcal{S}$ is nonempty, since $\mathcal{C} \in \mathcal{S}$, and it is partially ordered by inclusion. We check that the hypotheses of Zorn's lemma hold. Let $\mathcal{T}$ be a chain in $\mathcal{S}$, and let $\mathcal{A} = \bigcup \mathcal{T}$, the union of all sets in $\mathcal{T}$. Since each set in $\mathcal{T}$ is contained in $\mathcal{D}$ and contains $\mathcal{C}$, the same is true for $\mathcal{A}$. Therefore, $\mathcal{A} \in \mathcal{S}$. Moreover, $\mathcal{A}$ is clearly an upper bound for $\mathcal{T}$, since every set in $\mathcal{T}$ is contained in $\mathcal{A}$. By Zorn's lemma, there is a maximal element $\mathcal{B}$ of $\mathcal{S}$. We claim that $\mathcal{B}$ is a basis. Since $\mathcal{B} \in \mathcal{S}$, we see that $\mathcal{B}$ is linearly independent. To show that $\mathcal{B}$ spans V, let W be the span of $\mathcal{B}$. Since $\mathcal{D}$ spans V, it is sufficient to show that each $v \in \mathcal{D}$ is also in W. Suppose that there is a $v \in \mathcal{D}$ with $v \notin W$. Then v is not a linear combination of vectors in $\mathcal{B}$, so $\mathcal{B} \cup \{v\}$ is linearly independent. Moreover, $\mathcal{B} \cup \{v\} \subseteq \mathcal{D}$. However, this contradicts the maximality of $\mathcal{B}$; hence, $v \in W$. Therefore, $\mathcal{B}$ does span V, finishing the proof. □

Theorem 1.2 *Let V be an F-vector space. If $\mathcal{B}_1$ and $\mathcal{B}_2$ are bases for V, then $\mathcal{B}_1$ and $\mathcal{B}_2$ have the same cardinality.*

Proof. We prove only part of this theorem, taking for granted the following statement: If V is spanned by a finite set $\mathcal{D}$ and if $\mathcal{C}$ is any linearly independent set in V, then $|\mathcal{C}| \leq |\mathcal{D}|$. A proof of this fact is a standard

step in showing the uniqueness of the size of a basis for finite dimensional vector spaces.

Armed with this fact, we prove the theorem for infinite dimensional vector spaces. If one of $\mathcal{B}_1$ or $\mathcal{B}_2$ is finite, the fact above forces both to be finite. So, suppose that both are infinite. For each $v_i \in \mathcal{B}_2$, write $v_i = \sum_j \alpha_{ij} w_j$ with the $w_j \in \mathcal{B}_1$. Let $\mathcal{J}_i = \{w_j : \alpha_{ij} \neq 0\}$, a finite subset of $\mathcal{B}_1$. Let $\mathcal{K} = \bigcup_i \mathcal{J}_i$, a subset of $\mathcal{B}_1$. Since each element of $\mathcal{B}_2$ is a linear combination of elements of $\mathcal{K}$, the vector space V is spanned by $\mathcal{K}$. Since $\mathcal{K} \subseteq \mathcal{B}_1$ and $\mathcal{B}_1$ is a basis for V, this forces $\mathcal{K} = \mathcal{B}_1$. By Theorem 2.1 of Appendix B, $|\mathcal{K}| \leq \aleph_0 \, |\mathcal{B}_2|$, since $|\mathcal{J}_i| \leq \aleph_0$ for each i, and the union is over all elements of $\mathcal{B}_2$. But $\mathcal{B}_2$ is infinite; hence, $\aleph_0 \, |\mathcal{B}_2| = |\mathcal{B}_2|$. Therefore, $|\mathcal{B}_1| = |\mathcal{K}| \leq |\mathcal{B}_2|$. Reversing the roles of $\mathcal{B}_1$ and $\mathcal{B}_2$ gives the other inequality, proving that $|\mathcal{B}_1| = |\mathcal{B}_2|$. □

This theorem allows us to define the dimension of a vector space. The *dimension* of a vector space V is the cardinality of any basis of V. By the theorem, this is a well-defined invariant of the vector space. If V has a finite basis, then V is said to be a *finite dimensional* vector space.

2 Linear Transformations

Let V and W be vector spaces over a field F. A *linear transformation* from V to W is an F-vector space homomorphism from V to W. Let $\hom_F(V, W)$ be the set of all linear transformations from V to W. Then $\hom_F(V, W)$ is an F-vector space, where addition is defined by $(S + T)(v) = S(v) + T(v)$ and scalar multiplication by $(\alpha T)(v) = \alpha(T(v))$. It is straightforward to prove that $\hom_F(V, W)$ is indeed a vector space with these operations.

If $W = V$, then $\hom_F(V, V)$ can be given a multiplication. Define multiplication by $S \cdot T = S \circ T$, the usual function composition. It is not hard to show that $S \circ T$ is again a linear transformation and that $\hom_F(V, V)$ is an associative ring under these operations. We can give a more concrete description of this ring using bases. Suppose that V is a finite dimensional vector space and that $\{v_1, \ldots, v_n\}$ is a basis for V. Let $T \in \hom_F(V, V)$. Then $T(v_j) = \sum_i \alpha_{ij} v_i$ for some $\alpha_{ij} \in F$. Let $M(T)$ be the $n \times n$ matrix (α_{ij}). A straightforward calculation shows that

$$\begin{aligned} M(S + T) &= M(S) + M(T), \\ M(S \circ T) &= M(S) \cdot M(T), \\ M(\alpha T) &= \alpha M(T). \end{aligned}$$

Therefore, the map $\theta : T \mapsto M(T)$ is a ring and vector space homomorphism from $\hom_F(V, V)$ to $M_n(F)$, the ring of $n \times n$ matrices over F. Moreover, we see that θ is a bijection. To prove injectivity, suppose that $M(T)$ is the zero matrix. Then $T(v_j) = 0$ for each j. Since every element

of V is a linear combination of the v_j, this forces T to be the zero map. Therefore, θ is injective. To show that θ is surjective, take $(\alpha_{ij}) \in M_n(F)$. It is an easy calculation to show that the formula

$$S\left(\sum_j a_j v_j\right) = \sum_j a_j \left(\sum_i \alpha_{ij} v_i\right)$$

gives a well-defined linear transformation with $M(S) = (\alpha_{ij})$. This shows that θ is surjective. Therefore, $\hom_F(V,V) \cong M_n(F)$. In fact, if $u_1, \ldots, u_n$ is any collection of elements of V, then there is a uniquely determined linear transformation $\varphi : V \to V$ given by $\varphi(v_i) = u_i$. On a general element of V, the map φ is given by $\varphi(\sum_j a_j v_j) = \sum_j a_j u_j$. Thus, linear transformations can be described in terms of a basis. As a vector space, $\hom_F(V,V)$ has dimension n^2. This can be seen by showing that the set $\{e_{ij} : 1 \le i,j \le n\}$ of "matrix units" is a basis for $M_n(F)$, where e_{ij} is the matrix of zeros, except for a 1 in the (i,j) entry.

The isomorphism $\theta : \hom_F(V,V) \cong M_n(F)$ does depend on the choice of basis. Given another basis $\{w_j\}$ of V, we obtain another isomorphism $\phi : \hom_F(V,V) \cong M_n(F)$. How do these isomorphisms differ? Let $S : V \to V$ be the linear transformation given by $S(v_j) = w_j$, and let B be the matrix $M(S)$ calculated with respect to the basis $\{v_i\}$. If $T \in \hom_F(V,V)$, we write $M(T)_{\mathcal{V}}$ and $M(T)_{\mathcal{W}}$, respectively, for the matrices obtained from T by using the bases $\mathcal{V} = \{v_i\}$ and $\mathcal{W} = \{w_i\}$, respectively. A matrix calculation shows that

$$M(T)_{\mathcal{W}} = B^{-1} M(T)_{\mathcal{V}} B.$$

This relation between matrix representations of linear transformations using different bases allows us to define the determinant and trace of a linear transformation. Let $T \in \hom_F(V,V)$, and let $A = M(T)$ be the matrix representation of T with respect to some basis. Then we define the determinant and trace of T by $\det(T) = \det(A)$ and $\mathrm{Tr}(T) = \mathrm{Tr}(A)$, respectively. These definitions are well defined, since $\det(B^{-1}AB) = \det(A)$ and $\mathrm{Tr}(B^{-1}AB) = \mathrm{Tr}(A)$ for any invertible matrix B.

The final result we describe in this section is the Cayley–Hamilton theorem. Let $A \in M_n(F)$. The *characteristic polynomial* $\chi_A(x)$ of A is the polynomial $\det(xI - A)$, where I is the $n \times n$ identity matrix. This is a monic polynomial of degree n. For instance, if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then

$$\begin{aligned} \chi_A(x) &= x^2 - (a+d)x + (ad - bc) \\ &= x^2 - \mathrm{Tr}(A)x + \det(A). \end{aligned}$$

Since

$$\det(xI - B^{-1}AB) = \det(B^{-1}(xI - A)B) = \det(xI - A),$$

we can define the characteristic polynomial of a linear transformation by $\chi_T(x) = \chi_A(x)$ if A is any matrix representation of T.

Let $f(x) \in F[x]$, and write $f(x) = \sum_i a_i x^i$. We can evaluate f at A by setting $f(A) = \sum_i a_i A^i$, where $A^0 = I$. If A is an $n \times n$ matrix, then there is a nonzero polynomial f with $f(A) = 0$; to show the existence of such an f, the $n^2 + 1$ elements $I, A, \dots, A^{n^2}$ form a dependent set in $M_n(F)$, since this vector space has dimension n^2. Therefore, there are $\alpha_i \in F$ with $\sum_{i=0}^{n^2} \alpha_i A^i = 0$. Letting $f(x) = \sum_i \alpha_i x^i$ proves our claim. Given a matrix A, the *minimal polynomial* of A is the monic polynomial $p(x)$ of least degree such that $p(A) = 0$. The Cayley–Hamilton theorem relates the characteristic and minimal polynomials of a matrix.

Theorem 2.1 (Cayley–Hamilton) *Let A be an $n \times n$ matrix and $\chi_A(x)$ be the characteristic polynomial of A. Then $\chi_A(A) = 0$. Moreover, if $p(x)$ is the minimal polynomial of A, then $p(x)$ divides $\chi_A(x)$, and these two polynomials have the same irreducible divisors.*

Proof. A proof of this result can be found in most nonelementary books on linear algebra. We give a proof that uses the structure theorem for finitely generated modules over a PID and the rational canonical form. For a proof of this structure theorem and more information on this approach, see Chapter 5 of Walker [27]. Let $V = F^n$, an n-dimensional F-vector space. By using A, we can define an $F[x]$-module structure on V as follows: If $f(x) = \sum_{i=0}^{m} a_i x^i \in F[x]$, then define $f(x)v = \sum_{i=0}^{m} a_i A^i v$. We set $A^0 = I$ in order for this definition to make sense. It is elementary to show that V is an $F[x]$-module, and V is finitely generated as an $F[x]$-module, since it is generated as a module by a vector space basis. Therefore, there are elements $v_1, \dots, v_t \in V$ and polynomials $f_1, \dots, f_t \in F[x]$ such that

$$V = \bigoplus_{i=1}^{t} F[x]v_i \cong \bigoplus_{i=1}^{t} F[x]/(f_i).$$

Recall that $\operatorname{ann}(v_i) = \{f \in F[x] : fv_i = 0\}$ and that $\operatorname{ann}(v_i) = (f_i)$. Furthermore, we may assume that f_i divides f_{i+1} for each i. We will have proved the theorem once we verify that f_t is the minimal polynomial of A and that $f_1 \cdots f_t$ is the characteristic polynomial of A. From the description of $(f_i) = \operatorname{ann}(v_i)$, we see that $f_t v_i = 0$ for each i, so $f_t v = 0$ for all $v \in V$. By the definition of scalar multiplication, the nullspace of $f_t(A)$ is F^n, so $f_t(A) = 0$. Therefore, p divides f_t. For the reverse inclusion, since $p(A) = 0$, we see that $pv_t = 0$, so $p \in \operatorname{ann}(v_t) = (f_t)$. This gives the reverse divisibility, so $f_t = p$. This verifies our first claim. For the second, we use the rational canonical form of A. There is an invertible matrix B such that BAB^{-1} is the rational canonical form of A. The rational canonical form is

in block matrix form

$$\begin{pmatrix} C_1 & 0 & \cdots & 0 \\ 0 & C_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C_t \end{pmatrix},$$

where C_i is the companion matrix to f_i; if $f_i = x^s + \sum_{j<s} b_j x^j$, then

$$C_i = \begin{pmatrix} 0 & \cdots & 0 & -b_0 \\ 1 & 0 & \cdots & -b_1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & -b_{s-1} \end{pmatrix}.$$

Moreover, $\det(xI - C_i) = f_i$; this can be seen by expanding the determinant along the first row and using induction on $\deg(f_i)$. Thus,

$$\begin{aligned} \det(xI - A) &= \det(B(xI - A)B^{-1}) \\ &= \det(xI - BAB^{-1}) \\ &= f_1 \cdots f_t. \end{aligned}$$

This proves the second claim, so the theorem is proved. □

3 Systems of Linear Equations and Determinants

We give here a brief discussion of solving systems of linear equations. A system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m &= b_2, \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m &= b_n \end{aligned}$$

can be represented as a matrix equation $AX = B$, where $A = (a_{ij})$, $X = (x_i)$, and $B = (b_i)$. Multiplication by A determines a linear transformation $T : F^m \to F^n$. The existence of a solution is equivalent to the condition that B is in the image of T. The *rank* of A, denoted $\text{rank}(A)$, is the dimension of the image $\{Av : v \in F^m\}$ of T, a subspace of F^n. The rank of A is an integer no larger than $\min\{n, m\}$. If $\text{rank}(A) = n$, then the system above has a solution for every B. More generally, the image of T is spanned by the columns of A; hence, the image of T is the *column space* of A. Therefore, $\text{rank}(A)$ is equal to the dimension of the column space of A. A fundamental

fact about rank is that the rank of A is also equal to the dimension of the *row space* of A, the subspace of F^m spanned by the rows of A. For a proof of this, see Theorem 3.4.16 of [27].

Suppose that A is an $n \times n$ matrix. If $\det(A) \neq 0$, then A is an invertible matrix, and so the system $AX = B$ has a unique solution $X = A^{-1}B$ for any B. Therefore, $\text{rank}(A) = n$. If $\det(A) = 0$, then the system $AX = B$ cannot be solved for every B; to see this, suppose that there are X_i with $AX_i = e_i$, where $\{e_i\}$ is a basis for F^n. Then the matrix C whose ith column is X_i is an inverse of A; hence, $\det(A) \neq 0$, which is false. Therefore, $\text{rank}(A) < n$. Thus, the determinant function can help us to determine when square systems of linear equations can be solved.

4 Tensor Products

In Section 20, we make use of the tensor product of vector spaces. For readers unfamiliar with tensor products, we give the basics here. We only consider tensor products of vector spaces over a field, the only case that we need in Section 20. In order to work with tensor products, we need the concept of a bilinear map. Let U, V, and W be vector spaces over a field F. A *bilinear map* from $U \times V$ to W is a function $B : U \times V \to W$ such that

$$\begin{aligned} B(au_1 + bu_2, v) &= aB(u_1, v) + bB(u_2, v), \\ B(u, av_1 + bv_2) &= aB(u, v_1) + bB(u, v_2). \end{aligned}$$

for all scalars a, b, all $u, u_1, u_2 \in U$ and all $v, v_1, v_2 \in V$; that is, a bilinear map is linear in each component. To say this in another way, for all $u \in U$ and $v \in V$, the functions $B_u : V \to W$ and $B_v : U \to W$ given by

$$\begin{aligned} B_u(v) &= B(u, v), \\ B_v(u) &= B(u, v) \end{aligned}$$

are linear transformations.

The tensor product $U \otimes_F V$ can be defined as follows. Let M be the F-vector space with basis $\{(u, v) \in U \times V\}$; that is, for each pair (u, v) in $U \times V$, there is a corresponding basis vector in M. Let N be the subspace spanned by

$$\begin{aligned} &(au_1 + bu_2, v) - a(u_1, v) - b(u_2, v), \\ &(u, av_1 + bv_2) - a(u, v_1) - b(u, v_2), \\ &(au, v) - a(u, v), \\ &(u, av) - a(u, v) \end{aligned}$$

for all $a, b \in F$, all $u_1, u_2 \in U$, and all $v_1, v_2 \in V$. Then $U \otimes_F V$ is defined to be M/N. We will denote by $u \otimes v$ the coset $(u, v) + N$ in $U \otimes_F V$. Note

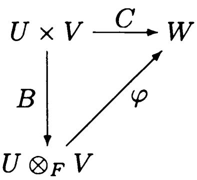
that since the (u, v) form a basis for M, each element of $U \otimes_F V$ is a sum of elements of the form $u \otimes v$. Looking at the generators of N, we obtain the following relations in $U \otimes_F V$:

$$\begin{aligned}
(au_1 + bu_2) \otimes v &= a(u_1 \otimes v) + b(u_2 \otimes v), \\
u \otimes (av_1 + bv_2) &= a(u \otimes v_1) + b(u \otimes v_2), \\
au \otimes v &= a(u \otimes v), \\
u \otimes av &= a(u \otimes v).
\end{aligned}$$

Define $B : U \times V \to U \otimes_F V$ by $B(u, v) = u \otimes v$. By the definition of tensor products, B is a bilinear map.

It is not terribly convenient to work with the construction of tensor products. The tensor product of U and V is best thought of in terms of the universal mapping property it satisfies.

Proposition 4.1 *Let U and V be F-vector spaces, and let $B : U \times V \to U \otimes_F V$ be the canonical bilinear map defined by $B(u, v) = u \otimes v$. If W is an F-vector space and $C : U \times V \to W$ is a bilinear map, then there is a unique linear transformation $\varphi : U \otimes_F V \to W$ such that $C = \varphi \circ B$; that is, the following diagram commutes:*



Proof. Let M and N be the vector spaces defined above in the construction of the tensor product. There is a unique linear transformation $f : M \to W$ with $f((u, v)) = C(u, v)$. The bilinearity of C implies precisely that the generators of N lie in $\ker(f)$. Thus, there is a linear transformation $\varphi : M/N \to W$ given by $\varphi((u, v) + N) = C(u, v)$. In other words, $\varphi(u \otimes v) = C(u, v)$. Since $B(u, v) = u \otimes v$, we see that $C = \varphi \circ B$. Moreover, this definition of φ is forced upon us by the restriction that $C = \varphi \circ B$; if $\sigma : U \otimes_F V \to W$ satisfies $C = \sigma \circ B$, then $\sigma(B(u, v)) = C(u, v)$, so $\sigma(u \otimes v) = C(u, v)$. Thus, σ and φ agree on the generators of $U \otimes_F V$, so $\sigma = \varphi$. □

Perhaps the most fundamental property of tensor products of vector spaces, other than the universal mapping property, is that the dimension of $U \otimes_F V$ is equal to $\dim_F(U) \cdot \dim_F(V)$. This is not a trivial fact to prove, which is the reason for the form of the next result.

Proposition 4.2 *Let U and V be finite dimensional F-vector spaces. Then $V \otimes_F \hom_F(U, F) \cong \hom_F(U, V)$. Consequently,* $\dim_F(U \otimes_F V) = \dim_F(U) \cdot \dim_F(V)$.

Proof. Define $C : V \times \hom_F(U, F) \to \hom_F(U, V)$ by

$$C(v, f)(u) = f(u)v.$$

We leave it to the reader to verify that $C(v, f)$ is indeed a linear transformation and that C is bilinear. By the universal mapping property, we get a linear transformation $\varphi : V \otimes_F \hom_F(U, F) \to \hom_F(U, V)$ given on generators by $\varphi(v \otimes f) = C(v, f)$.

Let $\{u_1, \ldots, u_n\}$ be a basis for U, and let $\{v_1, \ldots, v_m\}$ be a basis for V. Then the standard basis for $\hom_F(U, V)$ is $\{T_{ij}\}$, where

$$T_{ij}(u_k) = \begin{cases} v_j & \text{if } k = i \\ 0 & \text{if } k \neq i. \end{cases}$$

Taking the dual basis $\{\widehat{u_1}, \ldots, \widehat{u_n}\}$ for $\hom_F(U, F)$ (i.e., $\widehat{u_i}(u_j) = 0$ if $i \neq j$ and $\widehat{u_i}(u_i) = 1$), a short computation shows that $\varphi(v_i \otimes \widehat{u_j}) = T_{ij}$; hence, φ is surjective. Another short computation shows that $\{v_i \otimes \widehat{u_j}\}$ is a spanning set for $V \otimes_F \hom_F(U, F)$, which shows that $V \otimes_F \hom_F(U, F)$ has dimension at most nm, while the image of φ has dimension nm. Thus, φ is an isomorphism.

To finish the proof, we note that since U and $\hom_F(U, F)$ are isomorphic, the tensor products $V \otimes_F U$ and $V \otimes_F \hom_F(U, F)$ are isomorphic; hence, $V \otimes_F U$ has dimension nm. That $U \otimes_F V$ has the same dimension follows by reversing U and V and noting that $\hom_F(V, U)$ is also of dimension nm. □

Corollary 4.3 *Suppose that U and V are finite dimensional F-vector spaces. Let $\{u_1, \ldots, u_n\}$ and $\{v_1, \ldots, v_m\}$ be bases for U and V, respectively. Then $\{u_1 \otimes v_1, \ldots, u_n \otimes v_m\}$ is a basis for $U \otimes_F V$.*

Proof. The proof of the previous proposition shows that $\{v_i \otimes \widehat{u_j}\}$ is a basis for $V \otimes_F \hom_F(U, F)$. There is an isomorphism $\sigma : U \otimes_F V \to V \otimes_F \hom_F(U, F)$ given on generators by $\sigma(u \otimes v) \mapsto v \otimes \widehat{u}$ (see Problem 13), and this isomorphism sends $\{u_i \otimes v_j\}$ to the basis $\{v_i \otimes \widehat{u_j}\}$. This forces the set $\{u_i \otimes v_j\}$ to be a basis for $U \otimes_F V$. □

We will need to use tensor products of vector spaces of arbitrary dimension in Section 20. The following result is an analog of the previous corollary.

Proposition 4.4 *Let U and V be F-vector spaces. If $\{u_i\}_{i \in I}$ is a basis for U, then every element of $U \otimes_F V$ has a unique representation as a finite sum $\sum_i u_i \otimes v_i$ for some $v_i \in V$.*

Proof. If an element of $U \otimes_F V$ has two different representations in the form above, then subtracting the two yields an equation $\sum_{i=1}^n u_i \otimes v_i = 0$

with not all $v_i = 0$. By reducing the number of terms, if necessary, we may assume that the nonzero v_i in this equation are linearly independent. Let U_0 and V_0 be the subspaces of U and V generated by the u_i and the v_i, respectively. Extend $\{u_i\}$ and $\{v_i\}$ to bases of U and V, respectively. There are well-defined linear transformations $\sigma : U \to U_0$ and $\tau : V \to V_0$ with $\sigma(u_i) = u_i$ and $\tau(v_i) = v_i$ for $1 \le i \le n$, and all other u_j and v_j mapped to 0. The universal mapping property of tensor products shows that there is a linear transformation $\varphi : U \otimes_F V \to U_0 \otimes_F V_0$ given on generators by $\varphi(u \otimes v) = \sigma(u) \otimes \tau(v)$. Applying φ to the equation $\sum_{i=1}^n u_i \otimes v_i = 0$ yields the same equation in $U_0 \otimes_F V_0$, an impossibility by the previous corollary. This proves the proposition. □

We may ask why this proposition requires any proof at all, much less the roundabout proof given. The answer is that if we deal with modules over a ring R that is not a field, then it is common to have R-modules $M_0 \subseteq M$ and $N_0 \subseteq N$ such that $M_0 \otimes_R N_0$ is not isomorphic to the submodule of $M \otimes_R N$ consisting of elements of the form $\sum_i m_i \otimes n_i$ with $m_i \in M_0$ and $n_i \in N_0$. This pathological behavior happens quite frequently, even over rings such as $\mathbb{Z}$, although it does not occur with vector spaces over a field.

We finish this section by discussing the tensor product of F-algebras. If A is simultaneously a ring and an F-vector space, then A is called an *F-algebra* if

$$\alpha(ab) = (\alpha a)b = a(\alpha b)$$

for all $a, b \in A$ and all $\alpha \in F$; that is, there is a compatibility between the ring multiplication in A and the scalar multiplication. If A and B are F-algebras, then we can define a multiplication on $A \otimes_F B$ by the formula

$$\left(\sum_i a_i \otimes b_i\right)\left(\sum_i a_i' \otimes b_i'\right) = \sum_{i,j} a_i a_j' \otimes b_i b_j'. \tag{D.1}$$

On single tensors this says that $(a \otimes b)(a' \otimes b') = aa' \otimes bb'$. It needs to be checked that this formula gives a well-defined operation on $A \otimes_F B$. We leave it to the reader to verify the following result.

Proposition 4.5 *Let A and B be F-algebras. Then Equation* D.1 *is a well-defined multiplication on $A \otimes_F B$, and with respect to this multiplication, $A \otimes_F B$ is an F-algebra.*

Problems

1. Let V be an F-vector space. If $\mathcal{B}$ is a subset of V containing 0, show that $\mathcal{B}$ is linearly dependent.

2. Let $\mathcal{B}$ be a subset of a vector space V. Show that the set of all linear combinations of elements in $\mathcal{B}$ is a subspace of V.

3. Suppose that $\mathcal{C} \subseteq \mathcal{B}$ are bases of a vector space. Show that $\mathcal{C} = \mathcal{B}$.

4. Suppose that $\mathcal{C} \subseteq \mathcal{B}$ are subsets of a vector space V.

   (a) If $\mathcal{C}$ spans V, show that $\mathcal{B}$ also spans V.

   (b) If $\mathcal{B}$ is linearly independent, show that $\mathcal{C}$ is also linearly independent.

5. Show that the set of matrix units $\{e_{ij}\}$ described in Section 2 of this appendix is a basis for $M_n(F)$. More generally, show that the set of all $n \times m$ matrices over a field F is a vector space, and find the dimension by finding a basis analogous to that for $M_n(F)$.

6. If V and W are vector spaces of dimension n and m, respectively, show that $\hom_F(V, W)$ is isomorphic as a vector space to the space of all $n \times m$ matrices over F. Use this isomorphism and the previous problem to obtain a basis for $\hom_F(V, W)$.

7. Show that $\det(B^{-1}AB) = \det(A)$ and $\operatorname{Tr}(B^{-1}AB) = \operatorname{Tr}(A)$ for any matrix A and invertible matrix B.

8. Prove the equality $A(T)_{\mathcal{W}} = B^{-1}A(T)_{\mathcal{V}}B$ claimed in Section 2 of this appendix.

9. Find the characteristic polynomial of the following matrices.

   (a) $\begin{pmatrix} 2 & 3 \\ 4 & 2 \end{pmatrix}$,

   (b) $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$.

10. Find the minimal polynomial of the matrices in the previous problem.

11. Find the characteristic and minimal polynomials of $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$.

12. If U and V are F-vector spaces, show that $U \otimes_F V \cong V \otimes_F U$.

13. If U and V are finite dimensional F-vector spaces, show that there is an isomorphism between $U \otimes_F V$ and $V \otimes_F \hom_F(U, F)$ that sends $u \otimes v$ to $v \otimes \widehat{u}$, where $\widehat{u}$ is defined as follows: If $\{u_1, \ldots, u_n\}$ is a basis for U and if $\{\widehat{u_1}, \ldots, \widehat{u_n}\}$ is the dual basis for $\hom_F(U, F)$, if $u = \sum_i a_i u_i$, then $\widehat{u} = \sum_i a_i \widehat{u_i}$.

14. Let U, V, and W be F-vector spaces.

(a) Show that $U \otimes_F (V \oplus W) \cong (U \otimes_F V) \oplus (U \otimes_F W)$.

(b) Show that $U \otimes_F (V \otimes_F W) \cong (U \otimes_F V) \otimes_F W$.

15. Let U, V, and W be F-vector spaces. Show that

$$\hom_F(U \otimes_F V, W) \cong \hom_F(U, \hom_F(V, W)).$$

16. Give a proof of Proposition 4.5.

17. Let U and V be F-vector spaces with $\dim_F(U) < \infty$. If $\{u_1, \ldots, u_n\}$ is an F-basis for U, show that every element of $U \otimes_F V$ can be uniquely written in the form $\sum_i u_i \otimes v_i$ for some $v_i \in V$.

Appendix E
Topology

In Section 17 and in the sections that deal with algebraic geometry, we need to use some notions from topology. In this appendix, we give a brief description of these notions.

1 Topological Spaces

Let X be a set. A *topology* on X is a collection $\mathcal{T}$ of subsets of X that satisfy the following properties:

1. $X \in \mathcal{T}$ and $\varnothing \in \mathcal{T}$,
2. If $U, V \in \mathcal{T}$, then $U \cap V \in \mathcal{T}$,
3. If $\{U_i\}$ is a collection of subsets of X such that each $U_i \in \mathcal{T}$, then $\bigcup_i U_i \in \mathcal{T}$.

A set with a topology on it is called a *topological space*. The elements of a topology are called *open sets*. A subset C of X is called *closed* if $X - C$ is open. We can define a topology by specifying which are the closed sets. The closed sets of a topology on X satisfy the following properties.

1. Both X and $\varnothing$ are closed sets.
2. If A and B are closed sets, then $A \cup B$ is closed.
3. If $\{A_i\}$ is a collection of closed sets, then $\bigcap_i A_i$ is closed.

These properties follow immediately from the definition of a topology and the DeMorgan laws of set theory.

Example 1.1 The standard topology on $\mathbb{R}$ is defined as follows. A nonempty subset U of $\mathbb{R}$ is open, provided that for every $x \in U$ there is a positive number δ such that the open interval $(x-\delta, x+\delta)$ is contained in U. An easy exercise shows that this does make $\mathbb{R}$ into a topological space.

Example 1.2 Recall that a *metric space* is a set X together with a function d from $X \times X$ to the nonnegative real numbers such that (i) $d(x,x) = 0$ for all $x \in X$, and if $d(x,y) = 0$, then $x = y$, (ii) $d(x,y) = d(y,x)$ for all $x, y \in X$, and (iii) $d(x,y)+d(y,z) \geq d(x,z)$ for all $x, y, z \in X$. The function d is called a metric. We can use d to put a topology on X. A nonempty subset U of X is defined to be open, provided that for every $x \in U$ there is a positive number δ such that the *open ball*

$$B(x,\delta) = \{y \in X : d(x,y) < \delta\}$$

centered at x with radius δ is contained in U. This topology is called the *metric space topology*. The standard topology on $\mathbb{R}$ is an example of this construction. For another example, if $X = \mathbb{R}^n$, then we obtain a topology on $\mathbb{R}^n$, since we have a distance function on $\mathbb{R}^n$.

Example 1.3 If X is a topological space and Y is a subset of X, then we can put a topology on Y. We define a subset V of Y to be open if there is an open subset of X with $V = Y \cap U$. It is straightforward to show that Y is indeed a topological space. This topology on Y is called the *subspace topology*.

Example 1.4 Let X be a set. The *discrete topology* on X is the topology for which every subset of X is open.

Example 1.5 Let X be a set. We define a topology on X by defining a proper subset of X to be closed if it is finite. The definition of a topology is easy to verify in this case. Note that a nonempty subset is open exactly when its complement is finite. This topology is called the *finite complement topology* on X.

There are often more efficient ways to describe a topology than to list all of the closed sets. If X is a topological space, a *basis* for the topology on X is a collection of open subsets such that every open set is a union of elements from the basis. For example, the collection of open intervals forms a basis for the standard topology on $\mathbb{R}$. Similarly, the collection of open balls forms a basis for the metric topology on a metric space. A collection $\mathcal{C}$ of sets forms a basis for a topology on X provided that, given any two sets U and V in $\mathcal{C}$, for any $x \in U \cap V$ there is a set W in $\mathcal{C}$ such that $x \in W$ and $W \subseteq U \cap V$. The proof of this fact is left to Problem 1.

Example 1.6 Let R be a commutative ring, and let I be an ideal of R. The I-*adic topology* on R is defined as follows. A nonempty subset of R is open if it is the union of sets of the form $a + I^n$ for some $a \in R$ and $n \geq 0$. We set $I^0 = R$ for this definition. In other words, $\{a + I^n : a \in R, n \geq 0\}$ is a basis for this topology. The only nontrivial thing to verify to see that this does define a topology is that the intersection of two open sets is open. If $\bigcup_i (a_i + I^{n_i})$ and $\bigcup_j (b_j + I^{m_j})$ are open sets, then their intersection is $\bigcup_{i,j} (a_i + I^{n_i}) \cap (b_j + I^{m_j})$. It then suffices to show that $(a + I^n) \cap (b + I^m)$ is open for any $a, b \in R$ and $n, m \geq 0$. To prove this, we can assume that $n \geq m$, so $I^m \subseteq I^n$. If this intersection is empty, there is nothing to prove. If not, let $c \in (a + I^n) \cap (b + I^m)$. Then $c + I^n = a + I^n$ and $c + I^m = b + I^m$, so

$$\begin{aligned}(a + I^n) \cap (b + I^m) &= (c + I^n) \cap (c + I^m) \\ &= c + I^m,\end{aligned}$$

an open set.

Example 1.7 Here is an example that arises in algebraic geometry. Let R be a commutative ring, and let $X = \text{spec}(R)$ be the set of all prime ideals of R. If S is a subset of R, we set $Z(S) = \{P \in X : S \subseteq P\}$. We define the *Zariski topology* on X by defining a subset of X to be closed if it is of the form $Z(S)$ for some subset S of R. We verify that this is a topology on X. First, note that $R = Z(\{0\})$ and $\varnothing = Z(\{1\})$. Next, it is easy to see that $\bigcup_i Z(S_i) = Z(\bigcap_i S_i)$. Finally, we show that $Z(S) \cup Z(T) = Z(ST)$, where $ST = \{st : s \in S, t \in T\}$. Let $P \in Z(ST)$. If $P \notin Z(S)$, then there is an $s \in S$ with $s \notin P$. Since $st \in P$ for all $t \in T$, we see that $T \subseteq P$, since P is a prime ideal. Thus, $P \in Z(T)$. Therefore, $Z(ST) \subseteq Z(S) \cup Z(T)$. For the reverse inclusion, let $P \in Z(S) \cup Z(T)$. Then $S \subseteq P$ or $T \subseteq P$. Since P is an ideal, in either case we have $ST \subseteq P$, so $P \in Z(ST)$. We point out the relation between the Zariski topology on $\text{spec}(R)$ and the Zariski topology that we define in Section 21. We require some concepts from Section 21 in order to do this. Let C be an algebraically closed field, let V be a variety in C^n, and let $R = C[V]$ be the coordinate ring of V. Then V is homeomorphic to the subspace of $\text{spec}(R)$ consisting of all maximal ideals of R. This is mostly a consequence of the Nullstellensatz.

Example 1.8 Let X and Y be topological spaces. Then the product $X \times Y$ can be given a topology in the following way. We define a subset of $X \times Y$ to be open if it is a union of sets of the form $U \times V$, where U is an open subset of X and V is an open subset of Y; that is, the collection $\mathcal{C}$ of these subsets is a basis for the topology. It is easy to verify that this collection does satisfy the requirement to be a basis. If $(x, y) \in (U \times V) \cap (U' \times V')$, then $(U \cap U') \times (V \cap V')$ is a basic open set that contains (x, y) and is contained in $(U \times V) \cap (U' \times V')$. This topology on $X \times Y$ is called the

product topology. More generally, if $X_1, \dots, X_n$ is a collection of topological spaces, then we get a similar topology on $X_1 \times \cdots \times X_n$.

Example 1.9 Let I be a set, and let $\{X_i\}_{i \in I}$ be a collection of topological spaces. We can generalize the previous construction to define the product topology on $\prod_i X_i$. If I is infinite, then we need an extra step in the definition. Consider the set $\mathcal{S}$ of all subsets of $\prod_i X_i$ of the form $\prod_i U_i$, where U_i is open in X_i and $U_i = X_i$ for all but finitely many i. If I is finite, then $\mathcal{S}$ is the basis described in the previous example. If I is not finite, then we let $\mathcal{C}$ be the collection of all sets that are finite intersections of elements of $\mathcal{S}$. It is not hard to show that $\mathcal{C}$ does form a basis for a topology on $\prod_i X_i$, and we call this the product topology on $\prod_i X_i$. It is true that $\mathcal{S}$ also forms a basis for a topology on X, the *box topology*, but this topology is not as useful as the product topology.

2 Topological Properties

There are various properties of topological spaces that we need to discuss. Let X be a topological space. Then X is called *Hausdorff* if for every two distinct points $x, y \in X$, there are disjoint open sets U and V with $x \in U$ and $y \in V$. For example, if X is a metric space, then we see that the metric space topology is Hausdorff. If $x, y \in X$ are distinct points, let $\delta = \frac{1}{2}d(x,y)$. Then the open balls $B(x,\delta)$ and $B(y,\delta)$ are disjoint open sets containing x and y, respectively. The finite complement topology on an infinite set X is not Hausdorff, since any two nonempty open sets must have a nonempty intersection. If R is an integral domain, then we show that the Zariski topology on spec(R) is not Hausdorff either. We note that the zero ideal is prime and that $(0) \notin Z(S)$ for any S unless $Z(S) = \text{spec}(R)$. Consequently, (0) is contained in any nonempty open set. Therefore, any two nonempty open sets have a nonempty intersection, so spec(R) is not Hausdorff.

The next concept we discuss is compactness. If X is a topological space, then an *open cover* of X is a collection of open sets whose union is X. If $\{U_i\}$ is an open cover of X, then a finite subcover is a finite subset of the collection whose union is also X. The space X is called *compact* if every open cover of X has a finite subcover.

Example 2.1 The space $\mathbb{R}$ is not compact, since $\{(a, a+1) : a \in \mathbb{R}\}$ is an open cover of $\mathbb{R}$ that does not have a finite subcover. Subspaces of $\mathbb{R}^n$ may be compact. Recall that a subset Y of $\mathbb{R}^n$ is bounded if Y is contained in an open ball $B(0,\delta)$ for some δ. The Heine–Borel theorem says that a subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

Example 2.2 Let R be a commutative ring. The Zariski topology on spec(R) is compact, as we now show. Suppose that $\{U_i\}$ is an open cover of

spec(R). If $Z(S_i)$ is the complement of U_i, then $\bigcap_i Z(S_i) = Z(\bigcup S_i) = \varnothing$. We first point out that if I_i is the ideal generated by S_i, then $Z(I_i) = Z(S_i)$ and $Z(\bigcup S_i) = Z(\sum_i I_i)$. The ideal $\sum_i I_i$ cannot be a proper ideal, since if it is, then it is contained in a maximal ideal, and so $Z(\sum_i I_i) \neq \varnothing$. Thus, $\sum_i I_i = R$, so there is a finite subcollection $I_1, \ldots, I_n$ and elements $r_i \in I_i$ such that $r_1 + \cdots + r_n = 1$. Then $\sum_{i=1}^n I_i = R$, and so there is no prime ideal that contains each I_i. Consequently, $\bigcap_{i=1}^n Z(I_i) = \varnothing$, so $\bigcup_{i=1}^n U_i = \text{spec}(R)$. We have found a finite subcover of $\{U_i\}$, so spec(R) is compact.

Example 2.3 Let $\{X_i\}$ be a collection of compact topological spaces. Then the product $\prod_i X_i$ is compact in the product topology. This nontrivial fact is the Tychonoff theorem and can be found in Chapter 5 of Munkres [22].

Let X be a topological space, and let S be a subset of X. The *closure* $\overline{S}$ of S is defined to be the intersection of all closed sets that contain S. Since X is closed, the closure is a closed set that contains S. The main property about this concept is given in the following proposition. The simple proof is left to Problem 4.

Proposition 2.4 *Let X be a topological space, and let S be a subset of X.*

1. *If C is any closed set that contains S, then $\overline{S} \subseteq C$.*
2. *If U is an open set with $U \cap \overline{S} \neq \varnothing$, then $U \cap S \neq \varnothing$.*

One consequence of this proposition is that an element $x \in X$ is in the closure of a subset S, provided that for any open set U that contains x, we have $U \cap S \neq \varnothing$. This is a useful way to determine when an element is in $\overline{S}$.

If X is a topological space and Y is a subset of X, then Y is *dense* in X if $\overline{Y} = X$. For example, any set S is dense in its closure $\overline{S}$. The open interval $(0, 1)$ is dense in $[0, 1]$. If R is a commutative ring, then we show that any nonempty open subset of spec(R) is dense in spec(R). If U is an open set, then $\overline{U}$ is a closed subset of spec(R), and $U \cap (\text{spec}(R) - \overline{U}) = \varnothing$. However, we have seen that any two nonempty open sets in spec(R) have a nonempty intersection. This forces $\overline{U} = \text{spec}(R)$, so U is dense in spec(R).

We have not yet discussed functions between topological spaces. If X and Y are topological spaces, then a function $f : X \to Y$ is called *continuous* if $f^{-1}(V)$ is open in X for any open set V in Y. If X and Y are subsets of $\mathbb{R}$, then this definition of continuity is equivalent to the limit definition given in calculus; see Problem 6.

Let X be a topological space, and let $\sim$ be an equivalence relation on X. We let X^* be the set of equivalence classes, and for $x \in X$ we denote the equivalence class of x by $\overline{x}$. We have a natural surjective function $\pi : X \to$

X^* given by $f(x) = \overline{x}$. We define the *quotient topology* on X^* as follows. A subset Y of X^* is defined to be open if $\pi^{-1}(Y)$ is open in X. It is a simple exercise to show that this does define a topology on X^* and that π is continuous. Moreover, the quotient topology is the topology on X^* that has the fewest open sets for which π is continuous.

We end this appendix with a concept that will arise in Section 17. A topological space X is called *connected* if X is not the union of two disjoint closed sets. For example, $\mathbb{R}$ is a connected set, while the subspace $[0,1] \cup [2,3]$ is not connected. On the other extreme, a space X is called *totally disconnected* if the only connected subsets of X are singleton sets. A space with the discrete topology is totally disconnected. The topology on a Galois group we define in Section 17 is totally disconnected.

Problems

1. Let $\mathcal{C}$ be a collection of subsets of a set X such that for any $U, V \in \mathcal{C}$ and any $x \in U \cap V$ there is a $W \in \mathcal{C}$ such that $x \in W$ and $W \subseteq U \cap V$. By defining a subset of X to be open if it is a union of elements of $\mathcal{C}$, show that this gives a topology on X.

2. A topological space X is called *irreducible* if X is not the union of two proper closed subsets. If X is irreducible, show that every nonempty open subset of X is dense in X and that any two nonempty open sets have a nonempty intersection.

3. Let R be an integral domain. Show that $\text{spec}(R)$ is an irreducible space.

4. Prove Proposition 2.4.

5. Show that $\mathbb{Q}$ is a dense subset of $\mathbb{R}$ in the standard topology on $\mathbb{R}$.

6. Let X be an open interval in $\mathbb{R}$, and let $f : X \to \mathbb{R}$. Show that f is continuous according to the definition given above if and only if f is continuous according to the limit definition given in calculus.

7. Let R be a commutative ring, and let I be an ideal of R. Show that the I-adic topology on R is Hausdorff if and only if $\bigcap_{n=1}^{\infty} I^n = (0)$.

8. Let X and Y be topological spaces, and let $f : X \to Y$ be a continuous function. Define an equivalence relation $\sim$ on X by saying that $x \sim z$ if $f(x) = f(z)$. Prove that $\sim$ is an equivalence relation and that there is a continuous function $\overline{f} : X^* \to Y$ such that $\overline{f} \circ \pi = f$.

9. Let X be an infinite set, and put the finite complement topology on X. Prove that X is an irreducible space. Prove also that X is connected.

10. Let R and S be commutative rings, and let $f : R \to S$ be a ring homomorphism. We assume that $f(1) = 1$. If Q is a prime ideal of S, show that $f^{-1}(Q)$ is a prime ideal of R. Show that we have an induced map $f^* : \operatorname{spec}(S) \to \operatorname{spec}(R)$ and that this map is continuous with respect to the Zariski topology.

11. Let X be a topological space. Then X has the *finite intersection property* if for any collection $\{C_i\}$ of closed subsets, if the intersection of the C_i is empty, then there is a finite subcollection whose intersection is also empty. Prove that X has the finite intersection property if and only if X is compact.

12. Prove that $[0, 1]$ is a compact subspace of $\mathbb{R}$ without using the Heine–Borel theorem.

13. Prove the Heine–Borel theorem for $\mathbb{R}$.

14. Prove that $(0, 1)$ is not compact.

15. Prove that any interval in $\mathbb{R}$ is connected.

16. *The Cantor Set.* Let $X_1 = [0, 1]$. Remove the middle third $(1/3, 2/3)$ of this interval, and let X_2 be the resulting set. Remove the middle third of each of the two intervals that make up X_2, and let X_3 be the resulting set. If we continue this process, we obtain sets X_n for each positive integer n. Let $C = \bigcap_{n=1}^{\infty} X_n$. Prove that C is compact and totally disconnected, and that X does not contain any intervals.

References

[1] E. Artin, *Galois Theory*, University of Notre Dame Press, South Bend, IN, 1942.

[2] M. F. Atiyah and I. G. Macdonald, *Introduction to Commutative Algebra*, Addison-Wesley, Reading, MA, 1969.

[3] G. Cardano, *Ars Magna, or the Rules of Algebra*, Dover Publications Inc., New York, NY, 1993.

[4] C. Chevalley, *Introduction to the Theory of Algebraic Functions of One Variable*, Math. Surveys, No. 6, American Mathematical Society, Providence, RI, 1951.

[5] J.-L. Colliot-Thélène, Birational invariants, purity and the Gersten conjecture, pp. 333–358 in *K-Theory and Algebraic Geometry: Connections with Quadratic Forms and Division Algebras*, (eds. B. Jacob and A. Rosenberg), Proc. Symp. Pure Math., Vol. 58, Part 2, American Mathematical Society, Providence, R.I., 1995.

[6] P. K. Draxl, *Skew Fields*, Cambridge University Press, Cambridge, 1983.

[7] H. M. Edwards, *Galois Theory*, Springer-Verlag, New York, NY, 1984.

[8] H. B. Enderton, *Elements of Set Theory,* Academic Press, New York, NY, 1977.

[9] L. Gaal, *Classical Galois Theory*, Chelsea Pub. Co., New York, NY, 1971.

[10] D. J. H. Garling, *A Course in Galois Theory*, Cambridge University Press, Cambridge, 1986.

[11] R. Hartshorne, *Algebraic Geometry*, Springer-Verlag, New York, NY, 1977.

[12] I. N. Herstein, *Topics in Algebra*, 2nd ed., John Wiley & Sons, New York, NY, 1975.

[13] T. W. Hungerford, *Algebra*, Springer-Verlag, New York, NY, 1974.

[14] I. M. Isaacs, *Solution of polynomials by real radicals*, The American Mathematical Monthly (1985) Vol. 92, No. 8, 571–575.

[15] N. Jacobson, *Basic Algebra I*, 2nd ed., W. H. Freeman, New York, NY, 1985.

[16] N. Jacobson, *Basic Algebra II*, 2nd ed., W. H. Freeman, New York, NY, 1985.

[17] I. Kaplansky, *Fields and Rings*, 2nd ed., University of Chicago Press, Chicago, IL, 1965.

[18] L. Kappe and B. Warren, *An elementary test for the Galois group of a quartic polynomial*, The American Mathematical Monthly (1989) Vol. 96, No. 2, 133–137.

[19] E. Kunz, *Introduction to Commutative Algebra and Algebraic Geometry*, Birkhäuser, Boston, MA, 1985.

[20] P. J. McCarthy, *Algebraic Extensions of Fields*, Chelsea Pub. Co., New York, NY, 1966.

[21] D. Mead, *Newton's identities*, The American Mathematical Monthly (1992) Vol. 99, No. 8, 749–751.

[22] J. Munkres, *Topology, a first course*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1975.

[23] J. Rotman, *Galois Theory*, Universitext series, Springer-Verlag, New York, 1990.

[24] J.-P. Serre, *Cohomologie Galoisienne*, Lecture Notes in Mathematics, Vol. 5, Springer-Verlag, Berlin, 1964.

[25] S. S. Shatz, *Profinite Groups, Arithmetic, and Geometry*, Annals of Math. Studies, Vol. 67, Princeton University Press, Princeton, NJ, 1972.

[26] R. R. Stoll, *Set Theory and Logic,* Dover Publications Inc., New York, 1961.

[27] E. Walker, *Introduction to Abstract Algebra*, Random House, New York, NY, 1987.

[28] E. Weiss, *Algebraic Number Theory*, Chelsea Pub. Co., New York, NY, 1963.

[29] O. Zariski and P. Samuel, *Commutative Algebra Volume I*, Springer-Verlag, New York, NY, 1958.

[30] O. Zariski and P. Samuel, *Commutative Algebra Volume II*, Springer-Verlag, New York, NY, 1960.

Index

Graduate Texts in Mathematics

continued from page ii

61 WHITEHEAD. Elements of Homotopy Theory.
62 KARGAPOLOV/MERLZJAKOV. Fundamentals of the Theory of Groups.
63 BOLLOBAS. Graph Theory.
64 EDWARDS. Fourier Series. Vol. I. 2nd ed.
65 WELLS. Differential Analysis on Complex Manifolds. 2nd ed.
66 WATERHOUSE. Introduction to Affine Group Schemes.
67 SERRE. Local Fields.
68 WEIDMANN. Linear Operators in Hilbert Spaces.
69 LANG. Cyclotomic Fields II.
70 MASSEY. Singular Homology Theory.
71 FARKAS/KRA. Riemann Surfaces. 2nd ed.
72 STILLWELL. Classical Topology and Combinatorial Group Theory. 2nd ed.
73 HUNGERFORD. Algebra.
74 DAVENPORT. Multiplicative Number Theory. 2nd ed.
75 HOCHSCHILD. Basic Theory of Algebraic Groups and Lie Algebras.
76 IITAKA. Algebraic Geometry.
77 HECKE. Lectures on the Theory of Algebraic Numbers.
78 BURRIS/SANKAPPANAVAR. A Course in Universal Algebra.
79 WALTERS. An Introduction to Ergodic Theory.
80 ROBINSON. A Course in the Theory of Groups. 2nd ed.
81 FORSTER. Lectures on Riemann Surfaces.
82 BOTT/TU. Differential Forms in Algebraic Topology.
83 WASHINGTON. Introduction to Cyclotomic Fields.
84 IRELAND/ROSEN. A Classical Introduction to Modern Number Theory. 2nd ed.
85 EDWARDS. Fourier Series. Vol. II. 2nd ed.
86 VAN LINT. Introduction to Coding Theory. 2nd ed.
87 BROWN. Cohomology of Groups.
88 PIERCE. Associative Algebras.
89 LANG. Introduction to Algebraic and Abelian Functions. 2nd ed.
90 BRØNDSTED. An Introduction to Convex Polytopes.
91 BEARDON. On the Geometry of Discrete Groups.
92 DIESTEL. Sequences and Series in Banach Spaces.
93 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part I. 2nd ed.
94 WARNER. Foundations of Differentiable Manifolds and Lie Groups.
95 SHIRYAEV. Probability. 2nd ed.
96 CONWAY. A Course in Functional Analysis. 2nd ed.
97 KOBLITZ. Introduction to Elliptic Curves and Modular Forms. 2nd ed.
98 BRÖCKER/TOM DIECK. Representations of Compact Lie Groups.
99 GROVE/BENSON. Finite Reflection Groups. 2nd ed.
100 BERG/CHRISTENSEN/RESSEL. Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions.
101 EDWARDS. Galois Theory.
102 VARADARAJAN. Lie Groups, Lie Algebras and Their Representations.
103 LANG. Complex Analysis. 3rd ed.
104 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part II.
105 LANG. $SL_2(\mathbf{R})$.
106 SILVERMAN. The Arithmetic of Elliptic Curves.
107 OLVER. Applications of Lie Groups to Differential Equations. 2nd ed.
108 RANGE. Holomorphic Functions and Integral Representations in Several Complex Variables.
109 LEHTO. Univalent Functions and Teichmüller Spaces.
110 LANG. Algebraic Number Theory.
111 HUSEMÖLLER. Elliptic Curves.
112 LANG. Elliptic Functions.
113 KARATZAS/SHREVE. Brownian Motion and Stochastic Calculus. 2nd ed.
114 KOBLITZ. A Course in Number Theory and Cryptography. 2nd ed.
115 BERGER/GOSTIAUX. Differential Geometry: Manifolds, Curves, and Surfaces.
116 KELLEY/SRINIVASAN. Measure and Integral. Vol. I.
117 SERRE. Algebraic Groups and Class Fields.
118 PEDERSEN. Analysis Now.

119 ROTMAN. An Introduction to Algebraic Topology.
120 ZIEMER. Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation.
121 LANG. Cyclotomic Fields I and II. Combined 2nd ed.
122 REMMERT. Theory of Complex Functions. *Readings in Mathematics*
123 EBBINGHAUS/HERMES et al. Numbers. *Readings in Mathematics*
124 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part III.
125 BERENSTEIN/GAY. Complex Variables: An Introduction.
126 BOREL. Linear Algebraic Groups.
127 MASSEY. A Basic Course in Algebraic Topology.
128 RAUCH. Partial Differential Equations.
129 FULTON/HARRIS. Representation Theory: A First Course. *Readings in Mathematics*
130 DODSON/POSTON. Tensor Geometry.
131 LAM. A First Course in Noncommutative Rings.
132 BEARDON. Iteration of Rational Functions.
133 HARRIS. Algebraic Geometry: A First Course.
134 ROMAN. Coding and Information Theory.
135 ROMAN. Advanced Linear Algebra.
136 ADKINS/WEINTRAUB. Algebra: An Approach via Module Theory.
137 AXLER/BOURDON/RAMEY. Harmonic Function Theory.
138 COHEN. A Course in Computational Algebraic Number Theory.
139 BREDON. Topology and Geometry.
140 AUBIN. Optima and Equilibria. An Introduction to Nonlinear Analysis.
141 BECKER/WEISPFENNING/KREDEL. Gröbner Bases. A Computational Approach to Commutative Algebra.
142 LANG. Real and Functional Analysis. 3rd ed.
143 DOOB. Measure Theory.
144 DENNIS/FARB. Noncommutative Algebra.
145 VICK. Homology Theory. An Introduction to Algebraic Topology. 2nd ed.
146 BRIDGES. Computability: A Mathematical Sketchbook.
147 ROSENBERG. Algebraic K-Theory and Its Applications.
148 ROTMAN. An Introduction to the Theory of Groups. 4th ed.
149 RATCLIFFE. Foundations of Hyperbolic Manifolds.
150 EISENBUD. Commutative Algebra with a View Toward Algebraic Geometry.
151 SILVERMAN. Advanced Topics in the Arithmetic of Elliptic Curves.
152 ZIEGLER. Lectures on Polytopes.
153 FULTON. Algebraic Topology: A First Course.
154 BROWN/PEARCY. An Introduction to Analysis.
155 KASSEL. Quantum Groups.
156 KECHRIS. Classical Descriptive Set Theory.
157 MALLIAVIN. Integration and Probability.
158 ROMAN. Field Theory.
159 CONWAY. Functions of One Complex Variable II.
160 LANG. Differential and Riemannian Manifolds.
161 BORWEIN/ERDÉLYI. Polynomials and Polynomial Inequalities.
162 ALPERIN/BELL. Groups and Representations.
163 DIXON/MORTIMER. Permutation Groups.
164 NATHANSON. Additive Number Theory: The Classical Bases.
165 NATHANSON. Additive Number Theory: Inverse Problems and the Geometry of Sumsets.
166 SHARPE. Differential Geometry: Cartan's Generalization of Klein's Erlangen Programme.
167 MORANDI. Field and Galois Theory.